OM protein - protein search, using sw model

```
        Copyright   GenCore version 5.1.6
        Copyright (c) 1993 - 2003 Compugen Ltd.
```

Run on: November 5, 2003, 16:58:44 ; Search time 50 Seconds
(without alignments)
1244.416 Million cell updates/sec

Title: US-09-688-672A-14
Perfect score: 1969
Sequence: 1 MSRAFIIDPTISAIDGLYDL........RAPVEADAGGQKVLVRNVV 392

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 15726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

```
Database : A_Geneseq_19Jun03:*
  1: /SIDS17/gcgdata/geneseq/geneseqp-embl/AA1980.DAT:*
  2: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1981.DAT:*
  3: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1982.DAT:*
  4: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1983.DAT:*
  5: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1984.DAT:*
  6: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1985.DAT:*
  7: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1986.DAT:*
  8: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1987.DAT:*
  9: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1988.DAT:*
 10: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1989.DAT:*
 11: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1990.DAT:*
 12: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1991.DAT:*
 13: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1992.DAT:*
 14: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1993.DAT:*
 15: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1994.DAT:*
 16: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1995.DAT:*
 17: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1996.DAT:*
 18: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1997.DAT:*
 19: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1998.DAT:*
 20: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA1999.DAT:*
 21: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2000.DAT:*
 22: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2001.DAT:*
 23: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2002.DAT:*
 24: /SIDS1/gcgdata/geneseq/geneseqp-embl/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1969	100.0	392	20	AAW73765	M. tuberculosis an
2	1969	100.0	392	20	AAW73655	M. tuberculosis an
3	1969	100.0	392	22	AAU08226	Polypeptide encode
4	1969	100.0	392	22	AAU01882	M. tuberculosis an
5	1969	100.0	392	23	ABU05959	M. tuberculosis an
6	1969	100.0	392	23	AAE29714	Mycobacterium sp.
7	1969	100.0	392	23	AAE17578	Mycobacterium spec
8	1963	99.7	543	22	AAU01905	M. tuberculosis an
9	1947	98.9	408	22	AAU01886	M. tuberculosis an
10	1947	98.9	408	22	AAU01887	M. tuberculosis an
11	1713	87.0	788	22	AAU01903	M. tuberculosis an
12	1258	63.9	394	23	ABU05402	M. tuberculosis an
13	1174	59.6	242	22	AAU01884	M. tuberculosis an
14	1061	53.9	219	22	AAU01902	M. tuberculosis an
15	1055	53.6	744	22	AAU01902	M. tuberculosis an
16	1055	53.6	815	22	AAU01904	M. tuberculosis an
17	658	33.4	137	22	AAU01885	M. tuberculosis an
18	148.5	7.5	269	23	ABP38616	Staphylococcus epi
19		7.4	911	23	ABP26657	Streptococcus poly
20	140	7.1	911	23	ABP29744	Streptococcus poly
21	138.5	7.0	689	23	ABP65485	Bifidobacterium lo
22	138.5	7.0	7339	24	AAQ16358	Human translocated
23	138	7.0	898	18	AAW31853	Mycobacterium tube
24	134.5	6.8	528	22	AAB82611	Spider recombinant
25	134	6.8	318	21	ABB1229	Mycobacterium tube
26	133.5	6.8	712	21	AAB08630	Amino acid sequenc
27	133.5	6.8	730	21	AAB08631	Fusion protein com
28	133	6.8	1020	23	AAB18321	Collagen like prot
29	133	6.8	1077	14	AAR37751	CLP 3.1 monomer co
30	133	6.8	1077	17	AAR93261	CLP 3.1 polymer se
31	133	6.8	1077	19	AAW57657	Collagen-like poly
32	132.5	6.7	291	22	AAB82608	Spider recombinant
33	131.5	6.7	162	15	AAR57103	Carrot glycine-ric
34	130.5	6.6	660	20	AAY01303	Human tropoelastin
35	129.5	6.6	501	16	AAR76076	Mycobacterium anti
36	129.5	6.6	515	21	AAY69135	Amino acid sequenc
37	129.5	6.6	571	21	AAY69077	Amino acid sequenc
38	129.5	6.6	646	18	AAW27178	Nephila clavipes s
39	129.5	6.6	655	23	ABB76672	Protein related to
40	129.5	6.6	681	22	AAB82609	Spider recombinant
41	129.5	6.6	691	22	AAB82610	Spider recombinant
42	129.5	6.6	698	20	AAY01302	Human tropoelastin
43	129.5	6.6	698	21	AAY69069	Amino acid sequenc
44	129.5	6.6	730	19	AAW46315	Human elastin cont
45	129.5	6.6	730	23	AAO17360	Human elastin. Ho

ALIGNMENTS

```
RESULT 1
AAW73765
ID   AAW73765 standard; Protein; 392 AA.
XX
AC   AAW73765;
XX
DT   24-MAR-1999  (first entry)
XX
DE   M. tuberculosis antigen clone hTcc#1 protein sequence.
XX
KW   Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
KW   infection.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9853075-A2.
XX
PD   26-NOV-1998.
XX
PF   20-MAY-1998;   98WO-US10407.
XX
PR   05-MAY-1998;   98US-0073010.
PR   20-MAY-1997;   97US-0859381.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAW;
XX
DR   WPI; 1999-045314/04.
DR   N-PSDB; AAX01177.
XX
PT   Polypeptide comprising immunogenic Mycobacterium tuberculosis
```

PT   antigen - useful for immunisation against M. tuberculosis infection
PT   to treat or prevent tuberculosis, and in diagnosis of tuberculosis
XX
PS   Claim 2; Page 87-88; 100pp; English.
XX
CC   This sequence represents an immunogenic portion of a Mycobacterium
CC   tuberculosis antigen of the invention. The polypeptides are useful for
CC   immunotherapy to treat or prevent tuberculosis (especially in humans),
CC   e.g. they can be included with an acceptable carrier in pharmaceutical
CC   compositions or included in vaccines, and administered to induce
CC   protective immunity in a patient against M. tuberculosis. Tuberculosis is
CC   a chronic, infectious disease generally caused by M. tuberculosis
CC   infection, and if left untreated typically results in serious
CC   complications and death. Fusion proteins containing the antigen, or DNA
CC   molecules can similarly be included with an acceptable carrier in
CC   pharmaceutical compositions or in vaccines and administered as above. The
CC   polypeptides are also useful for diagnosis of tuberculosis, by contacting
CC   dermal cells with at least one polypeptide and detecting an immune
CC   response (especially induration) on the patient's skin. Inhibiting the
CC   spread of tuberculosis requires vaccination and accurate diagnosis, since
CC   antibiotic therapy may not be effective due to the existence of an
CC   asymptomatic but contagious stage and to patient non-compliance. The
CC   polypeptides overcome concerns of safety and efficacy of current
CC   vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
CC   lack of sensitivity and specificity of existing diagnostic techniques.
XX
SQ   Sequence 392 AA;

```
Query Match             100.0%;  Score 1969;  DB 20;  Length 392;
Best Local Similarity   100.0%;  Pred. No. 2e-163;
Matches  392;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120

Qy  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300

Qy  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360

Qy  361 SEGAAGTEDAERAPVEADAGGGQKVLVRNVV 392
         ||||||||||||||||||||||||||||||
Db  361 SEGAAGTEDAERAPVEADAGGGQKVLVRNVV 392
```

RESULT 2
AAW73655
ID   AAW73655 standard; Protein; 392 AA.
XX
AC   AAW73655;
XX
DT   24-MAR-1999 (first entry)
XX
DE   M. tuberculosis antigen clone hTcc#1 protein sequence.
XX
KW   Antigen; M.tuberculosis protein; immunotherapy; tuberculosis; diagnosis;
KW   infection.

PT   antigen - useful for immunisation against M. tuberculosis infection
PT   to treat or prevent tuberculosis, and in diagnosis of tuberculosis
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO9853076-A2.
XX
PD   26-NOV-1998.
XX
PF   20-MAY-1998; 98WO-US10514.
XX
PR   05-MAY-1998; 98US-0073009.
PR   20-MAY-1997; 97US-0858998.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Alderson MR, Campos-Neto A, Dillon DC, Skeiky YAW;
XX
DR   WPI; 1999-045315/04.
DR   N-PSDB; AAX01143.
XX
PT   New isolated Mycobacterium tuberculosis antigens - used to develop
PT   products for the prevention, treatment and diagnosis of tuberculosis
PT   infection
XX
PS   Claim 2; Page 88-89; 104pp; English.
XX
CC   This sequence represents an immunogenic portion of a Mycobacterium
CC   tuberculosis antigen of the invention. The polypeptides are useful for
CC   immunotherapy to treat or prevent tuberculosis (especially in humans),
CC   e.g. they can be included with an acceptable carrier in pharmaceutical
CC   compositions or included in vaccines, and administered to induce
CC   protective immunity in a patient against M. tuberculosis. Tuberculosis is
CC   a chronic, infectious disease generally caused by M. tuberculosis
CC   infection, and if left untreated typically results in serious
CC   complications and death. Fusion proteins containing the antigen, or DNA
CC   molecules can similarly be included with an acceptable carrier in
CC   pharmaceutical compositions or in vaccines and administered as above. The
CC   polypeptides are also useful for diagnosis of tuberculosis, by contacting
CC   dermal cells with at least one polypeptide and detecting an immune
CC   response (especially induration) on the patient's skin. Inhibiting the
CC   spread of tuberculosis requires vaccination and accurate diagnosis, since
CC   antibiotic therapy may not be effective due to the existence of an
CC   asymptomatic but contagious stage and to patient non-compliance. The
CC   polypeptides overcome concerns of safety and efficacy of current
CC   vaccination with live bacteria (usually Bacillus Calmette-Guerin) and
CC   lack of sensitivity and specificity of existing diagnostic techniques.
XX
SQ   Sequence 392 AA;

```
Query Match             100.0%;  Score 1969;  DB 20;  Length 392;
Best Local Similarity   100.0%;  Pred. No. 2e-163;
Matches  392;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120

Qy  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300

Qy  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360

Qy  361 SEGAAGTEDAERAPVEADAGGGQKVLVRNVV 392
         ||||||||||||||||||||||||||||||
Db  361 SEGAAGTEDAERAPVEADAGGGQKVLVRNVV 392
```

```
Db   301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
         |||||||||||||||||||||||||||||||
Db   361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392


RESULT 3
AAU08226
ID   AAU08226 standard; Protein; 392 AA.
XX
AC   AAU08226;
XX
DT   17-DEC-2001 (first entry)
XX
DE   Polypeptide encoded by Mycobacterium tuberculosis clone hTcc.
XX
KW   Tuberculosis; Mycobacterium infection; gene therapy; anti bacterial;
KW   immunostimulant; clone hTcc.
XX
OS   Mycobacterium tuberculosis.
XX
PN   WO20162893-A2.
XX
PD   30-AUG-2001.
XX
PF   26-FEB-2001; 2001WO-US05992.
XX
PR   25-FEB-2000; 2000US-0185037.
PR   08-AUG-2000; 2000US-0223828.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Campos-Neto A, Skeiky Y, Ovendale P, Jen S, Lodes M;
XX
DR   WPI; 2001-536638/59.
DR   N-PSDB; AAS12487.
XX
PT   An isolated polypeptide comprising a Mycobacterium antigen, e.g., from
PT   Mycobacterium tuberculosis, useful in a vaccine for enhancing an immune
PT   response to and inhibiting development of a Mycobacterium infection -
XX
PS   Example 1; Page 156-157; 161pp; English.
XX
CC   The present invention relates to the isolation of Mycobacterium
CC   tuberculosis antigen polypeptides (e.g. Tb224) and the nucleic acids
CC   encoding them. The invention describes compounds and methods for the
CC   diagnosis of tuberculosis or for inducing protective immunity against
CC   tuberculosis. The compounds comprise at least one immunogenic portion
CC   of one or more Mycobacterium proteins and nucleic acid molecules
CC   encoding such polypeptides. The Mycobacterium proteins and nucleic acid
CC   molecules encoding them can be used in diagnostic kits for the detection
CC   of Mycobacterium infection in patients and biological samples. The
CC   compounds of the invention and antibodies directed against the
CC   Mycobacterium proteins may be used in vaccines for immunisation against
CC   Mycobacterium infections. The nucleic acids encoding the Mycobacterium
CC   proteins may be used in gene therapy. The present sequence represents
CC   the polypeptide encoded by M. tuberculosis clone hTcc#1.
XX
SQ   Sequence 392 AA;

Query Match        100.0%;  Score 1969;  DB 22;  Length 392;
Best Local Similarity  100.0%;  Pred. No. 2e-163;
Matches  392;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEBLAAAFPGDGWLGSAA 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MSRAFIIDPTISAIDGLYDLLGIGIG-PNQGGILYSSLEYFEKALEBLAAAFPGDGWLGSAA 60

Qy    61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
```

```
Qy   121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy   181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy   241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGPGGLPSLAQVHAA 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGPGGLPSLAQVHAA 300

Qy   301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY 360
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY 360

Qy   361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
         |||||||||||||||||||||||||||||||
Db   361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392


RESULT 4
AAU01882
ID   AAU01882 standard; Protein; 392 AA.
XX
AC   AAU01882;
XX
DT   29-AUG-2001 (first entry)
XX
DE   M. tuberculosis antigen HTCC#1/Mtb40.
XX
KW   HTCC#1; Mtb40; antigen; vaccine; tuberculosis; AIDS;
KW   acquired immunodeficiency disease.
XX
OS   Mycobacterium tuberculosis.
XX
FH   Key             Location/Qualifiers
FT   Peptide         1..20
FT                   /label= Peptide 1
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         14..33
FT                   /label= Peptide 2
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         27..46
FT                   /label= Peptide 3
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         40..59
FT                   /label= Peptide 4
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         53..72
FT                   /label= Peptide 5
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         66..85
FT                   /label= Peptide 6
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         79..98
FT                   /label= Peptide 7
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         92..111
FT                   /label= Peptide 8
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         105..124
FT                   /label= Peptide 9
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         118..137
FT                   /label= Peptide 10
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         131..150
FT                   /label= Peptide 11
FT                   /note= "Used for T-cell epitope mapping"
FT   Peptide         144..163
FT                   /label= Peptide 12
```

FT	Peptide	157..176
		/label= Peptide_13
		/note= "Used for T-cell epitope mapping"
FT	Peptide	170..189
		/label= Peptide_14
		/note= "Used for T-cell epitope mapping"
FT	Peptide	183..202
		/label= Peptide_15
		/note= "Used for T-cell epitope mapping"
FT	Peptide	196..215
		/label= Peptide_16
		/note= "Used for T-cell epitope mapping"
FT	Peptide	209..228
		/label= Peptide_17
		/note= "Used for T-cell epitope mapping"
FT	Peptide	222..241
		/label= Peptide_18
		/note= "Used for T-cell epitope mapping"
FT	Peptide	235..254
		/label= Peptide_19
		/note= "Used for T-cell epitope mapping"
FT	Peptide	248..267
		/label= Peptide_20
		/note= "Used for T-cell epitope mapping"
FT	Peptide	261..280
		/label= Peptide_21
		/note= "Used for T-cell epitope mapping"
FT	Peptide	274..293
		/label= Peptide_22
		/note= "Used for T-cell epitope mapping"
FT	Peptide	287..306
		/label= Peptide_23
		/note= "Used for T-cell epitope mapping"
FT	Peptide	300..319
		/label= Peptide_24
		/note= "Used for T-cell epitope mapping"
FT	Peptide	313..332
		/label= Peptide_25
		/note= "Used for T-cell epitope mapping"
FT	Peptide	325..344
		/label= Peptide_26
		/note= "Used for T-cell epitope mapping"
FT	Peptide	339..358
		/label= Peptide_27
		/note= "Used for T-cell epitope mapping"
FT	Peptide	352..371
		/label= Peptide_28
		/note= "Used for T-cell epitope mapping"
FT	Peptide	365..384
		/label= Peptide_29
		/note= "Used for T-cell epitope mapping"
FT	Peptide	373..392
		/label= Peptide_30
		/note= "Used for T-cell epitope mapping"

XX
PN WO200124820-A1.
XX
PD 12-APR-2001.
XX
PF 10-OCT-2000; 2000WO-US28095.
XX
PR 07-OCT-1999; 99US-0158338.
PR 07-OCT-1999; 99US-0158425.
XX
PA (CORI-) CORIXA CORP.
XX
PI Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
DR WPI; 2001-290576/30.
DR N-PSDB; AAS03773.
XX
PT Vaccinating against Mycobacteria infections in mammals using fusion

---

PT		proteins comprising combinations of heterologous antigens -
XX		
PS		Example 2; Fig 6; 168pp; English.
XX		
CC		The sequence represents Mycobacterium tuberculosis HTCC#1 (also known
CC		as Mtb40), an M. tuberculosis antigen. Compositions comprising at least
CC		2 heterologous antigens, as a fusion protein, and vectors expressing the
CC		fusion proteins are used as vaccines to prophylactically immunise
CC		mammals (especially humans) against infection by Mycobacteria. The
CC		compositions contain at least 2 heterologous antigens that increase the
CC		serological sensitivity of individuals infected with tuberculosis, a
CC		disease frequently affecting patients with acquired immunodeficiency
CC		disease, AIDS.
XX		
SQ		Sequence 392 AA;

Query Match 100.0%; Score 1969; DB 22; Length 392;
Best Local Similarity 100.0%; Pred. No. 2e-163;
Matches 392; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy     1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAPPGDGWLGSAA 60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAPPGDGWLGSAA 60

Qy    61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120

Qy   121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy   181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy   241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAA 300

Qy   301 STRQALRPRADGPVGAAAEQVGGGQSQLVSAQGSQGMGGPVGMGGWMHPSSGASKGTTTKKY 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 STRQALRPRADGPVGAAAEQVGGGQSQLVSAQGSQGMGGPVGMGGWMHPSSGASKGTTTKKY 360

Qy   361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
           ||||||||||||||||||||||||||||||||
Db   361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
```

RESULT 5
ABU05959
ID ABU05959 standard; Protein; 392 AA.
XX
AC ABU05959;
XX
DT 08-APR-2003 (first entry)
XX
DE M. tuberculosis and M. leprae marker protein #610.
XX
KW Mycobacterioses; survival; virulence; protective antigen; vaccine;
KW mycobacterial disease; tuberculosis; leprosy.
XX
OS Mycobacterium tuberculosis.
OS Mycobacterium leprae.
XX
PN WO200274903-A2.
XX
PD 26-SEP-2002.
XX
PF 22-FEB-2002; 2002WO-IB01973.
XX
PR 22-FEB-2001; 2001US-270123P.
XX

PA   (INSP) INST PASTEUR.
XX
PI   Cole S;
XX
DR   WPI; 2002-759885/82.
XX
PT   Identifying and selecting genes for survival or virulence of
PT   mycobacteria by a comparative genomic analysis of the sequences of
PT   Mycobacterium tuberculosis and M. leprae -
XX
PS   Claim 17; Page 828; 874pp; English.
XX
CC   This invention relates to a novel method for identifying essential genes
CC   for survival or virulence of mycobacteria species. The method comprises
CC   aligning the genomic sequence of a first mycobacterium species on a
CC   genomic sequence of a second mycobacterium species and selecting a
CC   polynucleotide sequence that is highly conserved in both genomes with no
CC   counterparts in other bacterial genomic sequences and that corresponds
CC   to an essential gene for the survival or virulence of mycobacterium
CC   species. The method of the invention is useful for detecting M.
CC   tuberculosis or M. leprae infection. The method reduces the number of
CC   potential new targets and protective antigens for new drugs and vaccine
CC   compositions to treat and prevent mycobacterial diseases, particularly
CC   tuberculosis and leprosy. The present sequence represents a marker
CC   protein from Mycobacterium tuberculosis and Mycobacterium leprae
CC   identified using the method of the invention.
XX
SQ   Sequence   392 AA;

Query Match   100.0%;   Score 1969;   DB 23;   Length 392;
Best Local Similarity 100.0%; Pred. No.2e-163;
Matches 392; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DKYAGKNRNHVNFFQELADLDRQLISLIIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120

Qy  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300

Qy  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360

Qy  361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
         ||||||||||||||||||||||||||||||||
Db  361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
```

RESULT 6
AAE29714
ID   AAE29714 standard; Protein; 392 AA.
XX
AC   AAE29714;
XX
DT   27-JAN-2003 (first entry)
XX
DE   Mycobacterium sp. HTCC#1 antigenic protein.
XX
KW   Vaccine; immunity; diagnostic agent; gene therapy; HTCC#1 antigen;

KW   MTB40.
XX
OS   Mycobacterium sp.
XX
PN   WO200272792-A2.
XX
PD   19-SEP-2002.
XX
PF   13-MAR-2002; 2002WO-US08223.
XX
PR   13-MAR-2001; 2001US-275837P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky Y, Brannon M, Guderian J;
XX
DR   WPI; 2002-759844/82.
DR   N-PSDB; AAD47091.
XX
PT   New recombinant nucleic acid molecule comprising a Leishmania TSA,
PT   LeIF, M15 or 6H polynucleotide, useful as vaccine to elicit protective
PT   immunity against pathogenic microorganisms e.g. Leishmania and
PT   Mycobacterium tuberculosis   -
XX
PS   Disclosure; Page 103-104; 155pp; English.
XX
CC   The invention relates to a recombinant nucleic acid molecule encoding a
CC   fusion polypeptide. The recombinant nucleic acid comprises a heterologous
CC   polynucleotide sequence encoding an antigen or an antigenic fragment from
CC   Mycobacterium sp. and a Leishmania polynucleotide sequence encoding a
CC   polypeptide or its fragment. The Leishmania polynucleotide is selected
CC   from TSA, LeIF, M15, and 6H polynucleotides. Sequences of the invention
CC   are used in methods for eliciting immune response in mammals. They are
CC   useful as vaccines to elicit protective immunity against pathogenic
CC   microorganisms such as Leishmania and Mycobacterium tuberculosis. Fusion
CC   polypeptides are used for enhancing the expression of polynucleotides,
CC   as in vivo diagnostic agents and for raising antibodies in a non-human
CC   animal. The invention is used in gene therapy. The present sequence is
CC   Mycobacterium sp. HTCC#1 antigenic protein. HTCC#1 is also referred
CC   to as MTB40.
XX
SQ   Sequence   392 AA;

Query Match   100.0%;   Score 1969;   DB 23;   Length 392;
Best Local Similarity 100.0%; Pred. No.2e-163;
Matches 392; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

```
Qy    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DKYAGKNRNHVNFFQELADLDRQLISLIIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120

Qy  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300

Qy  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360

Qy  361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
```

```
Db     361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392

RESULT 7
AAE17578
ID   AAE17578 standard; Protein; 392 AA.
XX
AC   AAE17578;
XX
DT   22-APR-2002  (first entry)
XX
DE   Mycobacterium species MTB40 (HTCC #1) protein.
XX
KW   Fusion protein; antigen; serological sensitivity; immune response;
KW   tuberculosis; infection; vaccine; MTB40; HTCC #1 protein.
XX
OS   Mycobacterium sp.
XX
PN   WO200198460-A2.
XX
PD   27-DEC-2001.
XX
PF   20-JUN-2001; 2001WO-US19959.
XX
PR   20-JUN-2000; 2000US-0597796.
PR   01-FEB-2001; 2001US-265737P.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky Y, Reed S, Alderson M;
XX
DR   WPI; 2002-147798/19.
DR   N-PSDB; AAD28349.
XX
PT   Composition comprising MTB39 antigen and MTB32A antigen from
PT   Mycobacterium species, useful for eliciting immune response in a
PT   subject -
XX
PS   Claim 9; Page 120; Page 136pp; English.
XX
CC   The present invention relates to fusion proteins containing at least
CC   two Mycobacterium species antigens, nucleotides encoding them and
CC   compositions comprising such fusion proteins. The present invention
CC   particularly relates to nucleic acids encoding fusion proteins that
CC   include two or more individual M. tuberculosis antigens which increase
CC   the serological sensitivity of sera from individuals infected with
CC   tuberculosis and methods for their use in diagnosis, prevention and
CC   treatment of tuberculosis infection. Sequences of the invention are
CC   useful for eliciting an immune response in a mammal, e.g., human,
CC   immunised with BCG. They are useful in the diagnosis, treatment and
CC   prevention of Mycobacterium infection. The fusion proteins and the
CC   polynucleotides are useful as diagnostic tools in patients infected
CC   with Mycobacterium, in vitro and in vivo assays for detecting humoral
CC   antibodies or cell-mediated immunity against M. tuberculosis, for the
CC   diagnosis of an infection or monitoring of disease progression, as
CC   immunogens to generate or elicit a protective immune response in a
CC   patient and for raising anti-M. tuberculosis antibodies in a non-human
CC   animal. Sequences of the invention are also used as vaccines. MTB32A
CC   fusion proteins of the invention are useful as in vivo diagnostic agents
CC   for intradermal skin test. The present sequence is Mycobacterium species
CC   MTB40 (HTCC #1) protein.
XX
SQ   Sequence 392 AA;

  Query Match        100.0%;  Score 1969;  DB 23;  Length 392;
  Best Local Similarity  100.0%;
  Matches  392;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy     1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKAIEELAAAFPGDGWLGSAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKAIEELAAAFPGDGWLGSAA 60

Qy    61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
```

```
Db    61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLIVVKTLINATQLLKLLAKLAELVAAAIAD 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLIVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy   181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLITGWVTGLFSRGVNSLESFFAGVPGLT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLITGWVTGLFSRGVNSLESFFAGVPGLT 240

Qy   241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300

Qy   301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGWHPSSGASKGTTTKKY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGWHPSSGASKGTTTKKY 360

Qy   361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
          ||||||||||||||||||||||||||||||||
Db   361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392

RESULT 8
AAU01905
ID   AAU01905 standard; Protein; 543 AA.
XX
AC   AAU01905;
XX
DT   29-AUG-2001  (first entry)
XX
DE   M. tuberculosis antigen HTCC#1 fusion protein #4.
XX
KW   TbRA12-HTCC#1; antigen; vaccine; tuberculosis;
KW   AIDS; acquired immunodeficiency disease; His Tag.
XX
OS   Mycobacterium tuberculosis.
OS   Synthetic.
XX
FH   Key         Location/Qualifiers
FT   Binding-site  3..8
FT              /label= Histidine tag
FT              /note= "Nickel chelating region used to aid
FT              purification of the protein"
FT   Misc-difference 541
FT              /label=OTHER
FT              /note= "in frame STOP codon"
XX
PN   WO200124820-A1.
XX
PD   12-APR-2001.
XX
PF   10-OCT-2000; 2000WO-US28095.
XX
PR   07-OCT-1999;  99US-0158338.
PR   07-OCT-1999;  99US-0158425.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
DR   WPI; 2001-290576/30.
XX
PT   Vaccinating against Mycobacteria infections in mammals using fusion
PT   proteins comprising combinations of heterologous antigens -
XX
PS   Claim 17; Fig 8; 168pp; English.
XX
CC   The sequence represents Mycobacterium tuberculosis fusion protein,
CC   TbRA12-HTCC#1 and includes a His tag at the N-terminus to aid
CC   purification. Compositions comprising at least 2 heterologous
CC   antigens, as a fusion protein, and vectors expressing the fusion
```

```
CC    proteins are used as vaccines to prophylactically immunise mammals
CC    (especially humans) against infection by Mycobacteria. The compositions
CC    contain at least 2 heterologous antigens that increase the serological
CC    sensitivity of individuals infected with tuberculosis, a disease
CC    frequently affecting patients with acquired immunodeficiency disease,
CC    AIDS.
XX
SQ    Sequence 543 AA;

Query Match           99.7%;    Score 1963;  DB 22;  Length 543;
Best Local Similarity 99.7%;    Pred. No. 1e-162;
Matches 391;  Conservative   0;  Mismatches   1;  Indels   0;  Gaps   0;

Qy    1  MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  149  MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 208

Qy   61  DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  209  DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 268

Qy  121  YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  269  YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 328

Qy  181  IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  329  IISDVADIIKGILGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 388

Qy  241  GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  389  GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 448

Qy  301  STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGPVGMGGPVGMHPSSGASKGTTTKKY 360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  449  STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGPVGMGGPVGMHPSSGASKGTTTKKY 508

Qy  361  SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
            ||||||||||||||||||||||||||||||||
Db  509  SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 540

RESULT 9
AAU01886
ID   AAU01886 standard; Protein; 408 AA.
XX
AC   AAU01886;
XX
DT   29-AUG-2001 (first entry)
XX
DE   M. tuberculosis antigen HTCC#1(TM-1)
XX
KW   HTCC#1(TM-1); antigen; vaccine; tuberculosis; AIDS;
KW   acquired immunodeficiency disease; His Tag.
XX
OS   Mycobacterium tuberculosis.
XX
FH   Key            Location/Qualifiers
FT   Binding-site   1..8
FT                  /label= Histidine tag
FT                  /note= "Nickel chelating region used to aid
FT                  purification of the protein"
FT   Misc-difference 169..180
FT                  /note= "These residues are deleted to form the
FT                  HTCC#1(TM-1) protein sequence"
FT   Misc-difference 405
FT                  /label= OTHER
FT                  /note= "In frame STOP codon"
XX
PN   WO200124820-A1.
XX
PD   12-APR-2001.
XX
```

```
PF
XX
PR   07-OCT-1999;  99US-0158338.
PR   07-OCT-1999;  99US-0158425.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
DR   WPI; 2001-290576/30.
DR   N-PSDB; AAS03777.
XX
PT   Vaccinating against Mycobacteria infections in mammals using fusion
PT   proteins comprising combinations of heterologous antigens -
XX
PS   Example 2; Fig 9a; 168pp; English.
XX
CC   The sequence is Mycobacterium tuberculosis HTCC#1(TM-1)
CC   representing HTCC#1 (an M. tuberculosis antigen) lacking the first
CC   transmembrane domain and includes a His tag at the N-terminus to aid
CC   purification. Compositions comprising at least 2 heterologous antigens,
CC   as a fusion protein, and vectors expressing the fusion proteins are used
CC   as vaccines to prophylactically immunise mammals (especially humans)
CC   against infection by Mycobacteria. The compositions contain at least 2
CC   heterologous antigens that increase the serological sensitivity of
CC   individuals infected with tuberculosis, a disease frequently affecting
CC   patients with acquired immunodeficiency disease, AIDS.
XX
SQ   Sequence 408 AA;

Query Match           98.9%;    Score 1947;  DB 22;  Length 408;
Best Local Similarity 99.0%;    Pred. No. 1.7e-161;
Matches 392;  Conservative   0;  Mismatches   0;  Indels   4;  Gaps   2;

Qy    1  MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA  60
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9  MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA  68

Qy   61  DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69  DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 128

Qy  121  YIPVVGHALSAAFQAPFCAGAMAVVGGA--LAYLVVKTLINA-TQLLKLLAKLAELVAA 176
            ||||||||||||||||||||||||||||   |||||||||||  |||||||||||||||
Db  129  YIPVVGHALSAAFQAPFCAGAMAVVGGALKLAYLVVKTLINAKLTQLLKLLAKLAELVAA 188

Qy  177  AIADIISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGV 236
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  189  AIADIISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGV 248

Qy  237  PGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQ 296
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  249  PGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQ 308

Qy  297  VHAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGPVGMGGPVGMHPSSGASKGTT 356
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  309  VHAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGPVGMGGPVGMHPSSGASKGTT 368

Qy  357  TKKYSEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
            ||||||||||||||||||||||||||||||||||||
Db  369  TKKYSEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 404

RESULT 10
AAU01887
ID   AAU01887 standard; Protein; 408 AA.
XX
AC   AAU01887;
XX
DT   29-AUG-2001 (first entry)
XX
DE   M. tuberculosis antigen HTCC#1(TM-2).
XX
```

KW   HTCC#1(TM-2); antigen; vaccine; tuberculosis; AIDS;
KW   acquired immunodeficiency disease; His Tag.
XX
OS   Mycobacterium tuberculosis.
XX
FH   Key   Location/Qualifiers
FT   Binding-site   1..8
FT   /label= Histidine tag
FT   /note= "Nickel chelating region used to aid
FT   purification of the protein"
FT   Misc-difference 110..213
FT   /note= "These residues are deleted to form the
FT   HTCC#1(TM-2) protein sequence"
FT   Misc-difference 405
FT   /label= OTHER
FT   /note= "In frame STOP codon"
XX
PN   WO200124820-A1.
XX
PD   12-APR-2001.
XX
PF   10-OCT-2000; 2000WO-US28095.
XX
PR   07-OCT-1999; 99US-0158338.
PR   07-OCT-1999; 99US-0158425.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
DR   WPI; 2001-290576/30.
DR   N-PSDB; AAS03778.
XX
PT   Vaccinating against Mycobacteria infections in mammals using fusion
PT   proteins comprising combinations of heterologous antigens -
PS   Example 2; Fig 9d; 168pp; English.
XX
CC   The sequence is Mycobacterium tuberculosis HTCC#1(TM-1)
CC   representing HTCC#1 (an M. tuberculosis antigen) lacking all the
CC   transmembrane domains and includes a His tag at the N-terminus to aid
CC   purification. Compositions comprising at least 2 heterologous antigens,
CC   as a fusion protein, and vectors expressing the fusion proteins are used
CC   as vaccines to prophylactically immunise mammals (especially humans)
CC   against infection by Mycobacteria. The compositions contain at least 2
CC   heterologous antigens that increase the serological sensitivity of
CC   individuals infected with tuberculosis, a disease frequently affecting
CC   patients with acquired immunodeficiency disease, AIDS.
XX
SQ   Sequence 408 AA;

Query Match   98.9%; Score 1947; DB 22; Length 408;
Best Local Similarity 99.0%; Pred. No. 1.7e-161;
Matches 392; Conservative 0; Mismatches 0; Indels 4; Gaps 2;

Qy   1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYPEKALEELAAAFPGDGWLGSAA 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   9 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYPEKALEELAAAFPGDGWLGSAA 68

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRD--ILEGAKKGLEFVRPVAVD 118
        |||||||||||||||||||||||||||||||||||||||| |||||||||||||||||
Db   69 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDKLILEGAKKGLEFVRPVAVD 128

Qy 119 LTYIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAI 178
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 129 LTYIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAI 188

Qy 179 ADIISDVADIIKGTLGEVWEFITNA--LNGLKELWDKLTGWVTGLFSRGWSNLESFFAGV 236
        ||||||||||||||||||||||||| |||||||||||||||||||||||||||||||||
Db 189 ADIISDVADIIKGTLGEVWEFITNAKLLNGLKELWDKLTGWVTGLFSRGWSNLESFFAGV 248

Qy 237 PGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQ 296
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 249 PGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQ 308

Qy 297 VHAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTT 356
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 309 VHAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTT 368

Qy 357 TKKYSEGAAGTEDAERAPVEADAGGQKVLVRNVV 392
        |||||||||||||||||||||||||||||||||
Db 369 TKKYSEGAAGTEDAERAPVEADAGGQKVLVRNVV 404

RESULT 11
AAU01903
ID   AAU01903 standard; Protein; 788 AA.
XX
AC   AAU01903;
XX
DT   29-AUG-2001 (first entry)
XX
DE   M. tuberculosis antigen HTCC#1 fusion protein #2.
XX
KW   HTCC#1(1-149)-TbH9-HTCC#1(161-392); antigen; vaccine; tuberculosis;
KW   AIDS; acquired immunodeficiency disease; His Tag.
XX
OS   Mycobacterium tuberculosis.
OS   Synthetic.
XX
FH   Key   Location/Qualifiers
FT   Binding-site   3..8
FT   /label= Histidine tag
FT   /note= "Nickel chelating region used to aid
FT   purification of the protein"
FT   Misc-difference 785
FT   /label= OTHER
FT   /note= "In frame STOP codon"
XX
PN   WO200124820-A1.
XX
PD   12-APR-2001.
XX
PF   10-OCT-2000; 2000WO-US28095.
XX
PR   07-OCT-1999; 99US-0158338.
PR   07-OCT-1999; 99US-0158425.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
DR   WPI; 2001-290576/30.
XX
PT   Vaccinating against Mycobacteria infections in mammals using fusion
PT   proteins comprising combinations of heterologous antigens -
PS   Example 2; Fig 11; 168pp; English.
XX
CC   The sequence represents Mycobacterium tuberculosis fusion protein,
CC   HTCC#1(1-149)-TbH9-HTCC#1(161-392) and includes a His tag at the
CC   N-terminus to aid purification. Compositions comprising at least 2
CC   heterologous antigens, as a fusion protein, and vectors expressing the
CC   fusion proteins are used as vaccines to prophylactically immunise mammals
CC   (especially humans) against infection by Mycobacteria. The compositions
CC   contain at least 2 heterologous antigens that increase the serological
CC   sensitivity of individuals infected with tuberculosis, a disease
CC   frequently affecting patients with acquired immunodeficiency disease,
CC   AIDS.
XX
SQ   Sequence 788 AA;

Query Match   87.0%; Score 1713; DB 22; Length 788;
Best Local Similarity 49.4%; Pred. No. 1.1e-140;
Matches 383; Conservative 2; Mismatches 2; Indels 7; Gaps 2;

```
Qy      1  MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALELAAAFPGDGWLGSAA    60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      9  MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALELAAAFPGDGWLGSAA    68

Qy     61  DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT   120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     69  DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT   128

Qy    121  YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLV-----------------------     153
           ||||||||||||||||||||||||||||||  :|
Db    129  YIPVVGHALSAAFQAPFCAGAMAVVGGALKLMVDFGALPEINSARMYAGPGSASLVAAA   188

Qy    154  -----------------------------------------------------------   153
Db    189  QMWDSVASDLFSAASAFQSVVWGLTVGSWIGSSAGLMVAAVSPYVAWMSVTAGQAELTAA   248

Qy    154  -----------------------------------------------------------   153
Db    249  QVRVAAAAYETAYGLTVPPVIAENRAELMILIATNLLGQNTPAIAVNEAEYGEMWAQDA    308

Qy    154  -----------------------------------VKTLI-------------------   158
                                               ||:|
Db    309  AAMFGYAAATATATATLLPEEAPEMTSAGGLLEQAAAVEEASDTAAANQLMNNVPQALQ    368

Qy    159  -----------------------------------------------------------   158
Db    369  QLAQPTQGTTPSSKLGGLWKTVSPHRSPISNMVSMANNHMSMINSGVSMTNTLSSMLKGF   428

Qy    159  -----------------------------------------------------------   158
Db    429  APAAAAQAVQTAAQNGVRAMSSIGSSLGSSLGGGVAANLGRAASVGSLSVPQAWAAANQ    488

Qy    159  -----------------------------------------------------------   158
Db    489  AVTPAARALPLTSLITSAAEBRGPGQMLGGLPVGQMGARAGGGLSGVLRVPPRPVMPHSPA  548

Qy    159  --NATQLLKLLAKLAELVAAAIADIISDVADIIKGTLGEVWEFITNALNGLKELWDKLTG   216
             |||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    549  AGKLTQLLKLLAKLAELVAAAIADIISDVADIIKGILGEVWEFITNALNGLKELWDKLTG   608

Qy    217  WVTIGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASL  276
           ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609  WVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASL   668

Qy    277  PALAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGM   336
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    669  PALAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGM   728

Qy    337  GGPVGMGGMHPSSGASKGTTTKYSEGAAAGTEDAERAPVEADAGGQKVLVRNVV        392
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    729  GGPVGMGGMHPSSGASKGTTTKYSEGAAAGTEDAERAPVEADAGGQKVLVRNVV        784
```

RESULT 12
ABU05402
ID   ABU05402 standard; Protein; 394 AA.
XX
AC   ABU05402;
XX
DT   08-APR-2003 (first entry)
XX
DE   M. tuberculosis and M. leprae marker protein #53.
XX
KW   Mycobacterioses; survival; virulence; protective antigen; vaccine;
KW   mycobacterial disease; tuberculosis; leprosy.
XX
OS   Mycobacterium tuberculosis.
OS   Mycobacterium leprae.
XX
PN   WO200274903-A2.
XX
PD   26-SEP-2002.
XX

PF   22-FEB-2002; 2002WO-IB01973.
XX
PR   22-FEB-2001; 2001US-270123P.
XX
PA   (INSP) INST PASTEUR.
XX
PI   Cole S;
XX
DR   WPI; 2002-759885/82.
XX
PT   Identifying and selecting genes for survival or virulence of
PT   mycobacteria by a comparative genomic analysis of the sequences of
PT   Mycobacterium tuberculosis and M. leprae -
XX
PS   Claim 17; Page 224-225; 874pp; English.
XX
CC   This invention relates to a novel method for identifying essential genes
CC   for survival or virulence of mycobacteria species. The method comprises
CC   aligning the genomic sequence of a first mycobacterium species on a
CC   genomic sequence of a second mycobacterium species and selecting a
CC   polynucleotide sequence that is highly conserved in both genomes with no
CC   counterparts in other bacterial genomic sequences and that corresponds
CC   to an essential gene for the survival or virulence of mycobacterium
CC   species. The method of the invention is useful for detecting M.
CC   tuberculosis or M. leprae infection. The method reduces the number of
CC   potential new targets and protective antigens for new drugs and vaccine
CC   compositions to treat and prevent mycobacterial diseases, particularly
CC   tuberculosis and leprosy. The present sequence represents a marker
CC   protein from Mycobacterium tuberculosis and Mycobacterium leprae
CC   identified using the method of the invention.
XX
SQ   Sequence   394 AA;

Query Match   63.9%; Score 1258; DB 23; Length 394;
Best Local Similarity   62.7%; Pred. No. 2.4e-101;
Matches 247; Conservative   55; Mismatches   90; Indels   2; Gaps   1;

```
Qy      1  MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALELAAAFPGDGWLGSAA    60
           ||.||||||| .|| .. ||||.||..||.||||| .:|||||.|||||||||||||||
Db      1  MSGAFIIDPTLKAIEAWHALLGIGVPNDDGVLYSSLSFFEKALEHLAAAFPGDGWLGSAA   60

Qy     61  DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT   120
           |||||.||:..|.|||||.|.|.:|.|||:|||:||| .||..|||||..:|||||.:| |
Db     61  DKYAGQNRKRVDIFQELAELDKELIELIHNQANSVQTRGILDGAKKALLFVRPVAIDLN   120

Qy    121  YIPVVGHALSAAFQAPPCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD   180
           |||:.|. :.|||:. .| |. |..| |||.|...:|. |.|..:... ::..||.:||.||
Db    121  YIPLVGSVMSASIQAQACAAAMAAVSGGLAYLLVQTAIHTAKFVALLARLAHLLASAVAD   180

Qy    181  IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT   240
           :.||.|.||||.:. :.:.||..::.||.:..||. :||.||| :..||.||.|.||||:
Db    181  VVSDGVAIIKGIVDHLWHFIAGALTGLKDIVEKIIHWFFGLFSHWWSRLHSFFGGIPGLS   240

Qy    241  GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA   300
           |||||||||||||| .||:|.|||| .|.||.|.:|||.||:|.| ||.|.|||.||.|||
Db    241  GATSGLSQVTGLFGVPGLAGSSGLLSGESLLSTENLPSLAGVGAGLGLGSLPQLAQLHAA   300

Qy    301  STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGMHPSSGASK--GTTTK        358
           ||||. .|:..|... .|. ::.||||| ||| |||||| |:. |||..||:|
Db    301  STRQGTRSQAGVSAELSTEQFGGQQEPVSAGGSQGMGGSQGMGGMTPASTKSKKDERKKK   360

Qy    359  KYSEGAAAGTEDAERAPVEADAGGQKVLVRNVV                             392
           |||||||||:|||||||:|. ..||| ||.| ||
Db    361  KYSEGAAAGTDDAERAPIEVQSGGGKRALAQHVV                            394
```

RESULT 13
AAU01883
ID   AAU01883 standard; Protein; 242 AA.
XX
AC   AAU01883;
XX
DT   29-AUG-2001 (first entry)

```
AAU01884
ID   AAU01884 standard; Protein; 219 AA.
XX
AC   AAU01884;
XX
DT   29-AUG-2001  (first entry)
XX
DE   M. tuberculosis antigen HTCC#1(184-392).
XX
KW   HTCC#1(184-392); antigen; vaccine; tuberculosis; AIDS;
KW   acquired immunodeficiency disease; His Tag.
XX
OS   Mycobacterium tuberculosis.
XX
FH   Key             Location/Qualifiers
FT   Binding-site    2..7
FT                   /label= Histidine tag
FT                   /note= "Nickel chelating region used to aid
FT                   purification of the protein"
FT   Misc-difference 216
FT                   /label= OTHER
FT                   /note= "In frame STOP codon"
XX
PN   WO20124820-A1.
XX
PD   12-APR-2001.
XX
PF   10-OCT-2000; 2000WO-US28095.
XX
PR   07-OCT-1999;  99US-0158338.
PR   07-OCT-1999;  99US-0158425.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky Y,  Reed S,  Houghton RL,  McNeill PD,  Dillon DC,  Lodes ML;
XX
DR   WPI; 2001-290576/30.
XX
PT   Vaccinating against Mycobacteria infections in mammals using fusion
PT   proteins comprising combinations of heterologous antigens -
XX
PS   Example 2; Fig 7b; 168pp; English.
XX
CC   The sequence is Mycobacterium tuberculosis HTCC#1(184-392) (a M.
CC   representing the two C-terminal transmembrane domains of HTCC#1 (an M.
CC   tuberculosis antigen) and includes a His tag at the N-terminus to aid
CC   purification. Compositions comprising at least 2 heterologous antigens,
CC   as a fusion protein, and vectors expressing the fusion proteins are used
CC   as vaccines to prophylactically immunise mammals (especially humans)
CC   against infection by Mycobacteria. The compositions contain at least 2
CC   heterologous antigens that increase the serological sensitivity of
CC   individuals infected with tuberculosis, a disease frequently affecting
CC   patients with acquired immunodeficiency disease, AIDS.
XX
SQ   Sequence   219 AA;

Query Match             53.9%;  Score 1061;  DB 22;  Length 219;
Best Local Similarity  100.0%;  Pred. No. 1.7e-84;
Matches  209;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy     184 DVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLTGAT 243
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8 DVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLTGAT 67

Qy     244 SGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAASTR 303
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68 SGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAASTR 127

Qy     304 QALRPRADGPVGAAAEQVGGGSQQLVSAQGSQCMGGGPVGMGGMPSSGASKGTTTKKYSEG 363
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 QALRPRADGPVGAAAEQVGGGSQQLVSAQGSQCMGGGPVGMGGMPSSGASKGTTTKKYSEG 187

Qy     364 AARGTEDAERAPVEADAGGGQKVLVRNVV 392
```

```
XX
DE   M. tuberculosis antigen HTCC#1(1-223).
XX
KW   HTCC#1(1-223); antigen; vaccine; tuberculosis; AIDS;
KW   acquired immunodeficiency disease; His Tag.
XX
OS   Mycobacterium tuberculosis.
XX
FH   Key             Location/Qualifiers
FT   Binding-site    2..7
FT                   /label= Histidine tag
FT                   /note= "Nickel chelating region used to aid
FT                   purification of the protein"
FT   Misc-difference 240
FT                   /label= OTHER
FT                   /note= "In frame STOP codon"
XX
PN   WO20124820-A1.
XX
PD   12-APR-2001.
XX
PF   10-OCT-2000; 2000WO-US28095.
XX
PR   07-OCT-1999;  99US-0158338.
PR   07-OCT-1999;  99US-0158425.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky Y,  Reed S,  Houghton RL,  McNeill PD,  Dillon DC,  Lodes ML;
XX
DR   WPI; 2001-290576/30.
DR   N-PSDB; AAS03774.
XX
PT   Vaccinating against Mycobacteria infections in mammals using fusion
PT   proteins comprising combinations of heterologous antigens -
XX
PS   Example 2; Fig 7a; 168pp; English.
XX
CC   The sequence is Mycobacterium tuberculosis HTCC#1(1-223),
CC   representing the first transmembrane domain of HTCC#1 (an M.
CC   tuberculosis antigen) and includes a His tag at the N-terminus to aid
CC   purification. Compositions comprising at least 2 heterologous antigens,
CC   as a fusion protein, and vectors expressing the fusion proteins are used
CC   as vaccines to prophylactically immunise mammals (especially humans)
CC   against infection by Mycobacteria. The compositions contain at least 2
CC   heterologous antigens that increase the serological sensitivity of
CC   individuals infected with tuberculosis, a disease frequently affecting
CC   patients with acquired immunodeficiency disease, AIDS.
XX
SQ   Sequence   242 AA;

Query Match             59.6%;  Score 1174;  DB 22;  Length 242;
Best Local Similarity  100.0%;  Pred. No. 2.7e-94;
Matches  232;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy       1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALELAAAFPGDGWLGSAA 60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALELAAAFPGDGWLGSAA 67

Qy      61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      68 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 127

Qy     121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLIAKLAELVAAAIAD 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     128 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLIAKLAELVAAAIAD 187

Qy     181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESF 232
               ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     188 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESF 239
```

RESULT 14.

```
Db     188  AAAGTEDAERAPVEADAGGGQKVLVRNVV 216

RESULT 15
AAU01902
ID   AAU01902 standard; Protein; 744 AA.
XX
AC   AAU01902;
XX
DT   29-AUG-2001  (first entry)
XX
DE   M. tuberculosis antigen HTCC#1 fusion protein #1.
XX
KW   HTCC#1(184-392)-TbH9-HTCC#1(1-129); antigen; vaccine; tuberculosis;
KW   AIDS; acquired immunodeficiency disease; His Tag.
XX
OS   Mycobacterium tuberculosis.
OS   Synthetic.
XX
FH   Key            Location/Qualifiers
FT   Binding-site   3..8
FT                  /label= Histidine_tag
FT                  /note= "Nickel chelating region used to aid
FT                  purification of the protein"
FT   Misc-difference 742
FT                  /label= OTHER
FT                  /note= "In frame STOP codon"
XX
PN   WO200124820-A1.
XX
PD   12-APR-2001.
XX
PF   10-OCT-2000; 2000WO-US28095.
XX
PR   07-OCT-1999; 99US-0158338.
PR   07-OCT-1999; 99US-0158425.
XX
PA   (CORI-) CORIXA CORP.
XX
PI   Skeiky Y, Reed S, Houghton RL, McNeill PD, Dillon DC, Lodes ML;
XX
DR   WPI; 2001-290576/30.
DR   N-PSDB; AAS03790.
XX
PT   Vaccinating against Mycobacteria infections in mammals using fusion
PT   proteins comprising combinations of heterologous antigens -
XX
PS   Claim 14; Fig 10; 168pp; English.
XX
CC   The sequence represents Mycobacterium tuberculosis fusion protein,
CC   HTCC#1(184-392)-TbH9-HTCC#1(1-129) and includes a His tag at the
CC   N-terminus to aid purification. Compositions comprising at least 2
CC   heterologous antigens, as a fusion protein, and vectors expressing the
CC   fusion proteins are used as vaccines to prophylactically immunise mammals
CC   (especially humans) against infection by Mycobacteria. The compositions
CC   contain at least 2 heterologous antigens that increase the serological
CC   sensitivity of individuals infected with tuberculosis, a disease
CC   frequently affecting patients with acquired immunodeficiency disease,
CC   AIDS.
XX
SQ   Sequence  744 AA;

Query Match      53.6%;  Score 1055;  DB 22;  Length 744;
Best Local Similarity  99.5%;  Pred. No. 3e-83;
Matches  208;  Conservative  0;  Mismatches  1;  Indels  0;  Gaps  0;

Qy  184  DVADIIKGTLGEVWEFITNALNGLKELWDKLITGWVTGLFSRGWSNLESFFAGVPGLTGAT 243
         ||||||||:||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    9  DVADIIKGILGEVWEFITNALNGLKELWDKLITGWVTGLFSRGWSNLESFFAGVPGLTGAT 68

Qy  244  SGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSFGGLPSLAQVHAASTR 303
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   69  SGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSFGGLPSLAQVHAASTR 128

Qy  304  QALRPRADGPVGAAAEQVGGQSQLVSAQSQGMGGPVGMGGMHPSSGASKGTTTKKYSEG 363
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  129  QALRPRADGPVGAAAEQVGGQSQLVSAQSQGMGGPVGMGGMHPSSGASKGTTTKKYSEG 188

Qy  364  AAAGTEDAERAPVEADAGGGQKVLVRNVV 392
         |||||||||||||||||||||||||||||
Db  189  AAAGTEDAERAPVEADAGGGQKVLVRNVV 217

Search completed: November  5, 2003, 16:59:50
Job time : 53 secs
```

This Page Blank (uspto)

```
              GenCore version 5.1.6
     Copyright (c) 1993 - 2003  Compugen Ltd.

OM protein - protein search, using sw model

Run on:     November  5, 2003, 17:00:27 ; Search time 21 Seconds
                                          (without alignments)
                                    789.803 Million cell updates/sec

Title:          US-09-688-672A-14
Perfect score:  1969
Sequence:       1 MSRAFIIDPTISAIDGLYDL.........RAPVEADAGGGQKVLVRNVV 392

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters:    328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :  Issued Patents AA:*
          1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
          2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
          3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
          4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
          5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
          6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.
```

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1969	100.0	392	4	US-09-073-009-138	Sequence 138, App
2	148.5	7.5	269	4	US-09-134-001C-1461	Sequence 3461, Ap
3	135	6.9	318	3	US-09-060-756-727	Sequence 727, App
4	135	6.9	318	4	US-09-670-314-727	Sequence 727, Appl
5	133.5	6.8	731	4	US-09-340-736E-1	Sequence 1, Appl
6	133	6.8	1077	1	US-07-972-032-82	Sequence 82, Appl
7	133	6.8	1077	2	US-08-642-255-95	Sequence 95, Appl
8	129.5	6.6	731	2	US-08-911-364-1	Sequence 1, Appl
9	129.5	6.6	733	3	US-08-464-700-2	Sequence 2, Appli
10	129.5	6.6	1271	2	US-08-095-734-2	Sequence 2, Appli
11	129.5	6.6	1271	3	US-08-444-623-2	Sequence 2, Appli
12	129.5	6.6	1271	3	US-08-471-869-2	Sequence 2, Appli
13	129.5	6.6	1271	4	US-09-342-563-2	Sequence 2, Appli
14	129.5	6.6	1271	5	PCT-US94-08267-2	Sequence 2, Appli
15	128.5	6.5	832	1	US-08-209-747-2	Sequence 2, Appli
16	128.5	6.5	832	4	US-08-458-298-2	Sequence 2, Appli
17	126.5	6.4	334	3	US-09-060-756-728	Sequence 728, App
18	126.5	6.4	334	4	US-09-670-314-728	Sequence 728, App
19	126	6.4	1958	1	US-07-945-283-2	Sequence 2, Appli
20	125	6.3	738	3	US-08-864-038A-3	Sequence 3, Appli
21	125	6.3	2088	4	US-09-548-372D-13	Sequence 13, Appl
22	125	6.3	2088	4	US-09-548-367D-13	Sequence 13, Appl
23	125	6.3	2088	4	US-09-551-853D-13	Sequence 13, Appl
24	124.5	6.3	1415	4	US-09-252-991A-26438	Sequence 26438, A
25	124	6.3	792	2	US-08-678-039A-40	Sequence 40, Appl
26	124	6.3	1056	1	US-08-212-237-6	Sequence 6, Appl
27	124	6.3	1056	4	US-08-806-029-29	Sequence 29, Appl
28	124	6.3	1056	5	PCT-US95-02772-6	Sequence 6, Appli
29	124	6.3	1136	4	US-08-806-029-9	Sequence 9, Appli
30	123.5	6.3	651	3	US-08-556-978B-19	Sequence 19, Appl
31	123.5	6.3	651	1	US-08-247-806-1	Sequence 1, Appli
32	123.5	6.3	718	1	US-08-425-069-2	Sequence 2, Appli
33	123.5	6.3	718	2	US-08-317-844B-2	Sequence 2, Appli
34	123.5	6.3	747	3	US-09-034-177-3	Sequence 3, Appli
35	122	6.2	159	1	US-07-609-716-104	Sequence 104, App
36	122	6.2	159	4	US-08-475-411A-104	Sequence 104, App
37	122	6.2	159	3	US-08-478-029A-104	Sequence 104, App
38	121.5	6.2	1694	1	US-08-494-168-2	Sequence 2, Appli
39	121	6.1	785	3	US-09-252-991A-32952	Sequence 32952, A
40	120	6.1	604	3	US-08-556-978B-63	Sequence 63, Appl
41	120	6.1	766	1	US-08-175-155-53	Sequence 53, Appl
42	120	6.1	766	3	US-08-477-509B-88	Sequence 88, Appl
43	120	6.1	766	2	US-08-707-237A-61	Sequence 61, Appl
44	120	6.1	766	3	US-08-482-085B-88	Sequence 88, Appl
45	120	6.1	766	4	US-09-444-791A-88	Sequence 88, Appl

ALIGNMENTS

```
RESULT 1
US-09-073-009-138
; Sequence 138, Application US/09073009
; Patent No. 655653
; GENERAL INFORMATION:
;  APPLICANT: Alderson, Mark
;  APPLICANT: Dillon, Davin C.
;  APPLICANT: Skeiky, Yasir A.W.
;  APPLICANT: Campos-Neto, Antonio
;  TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF TUBERCULOSIS AND ME
;  NUMBER OF SEQUENCES: 144
;  CORRESPONDENCE ADDRESS:
;  ADDRESSEE: SEED and BERRY
;  STREET: 6300 Coumbia Center, 701 Fifth Ave.
;  CITY: Seattle
;  STATE: Washington
;  COUNTRY: US
;  ZIP: 98104
;  COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/09/073,009
;  FILING DATE: 05-MAY-1998
;  CLASSIFICATION:
;  ATTORNEY/AGENT INFORMATION:
;  NAME: Maki, David J.
;  REGISTRATION NUMBER: 31,392
;  REFERENCE/DOCKET NUMBER: 210121.441C1
;  TELECOMMUNICATION INFORMATION:
;  TELEPHONE: 206-622-4900
;  TELEFAX:  206-682-6031
;  INFORMATION FOR SEQ ID NO: 138:
;  SEQUENCE CHARACTERISTICS:
;  LENGTH: 392 amino acids
;  TYPE: amino acid
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: protein
US-09-073-009-138

Query Match             100.0%;  Score 1969;  DB 4;  Length 392;
Best Local Similarity   100.0%;  Pred. No. 1.3e-175;
Matches  392;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60
```

```
Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTRDILEGAKKGLEFVRPVAVDLT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTRDILEGAKKGLEFVRPVAVDLT 120

Qy  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLIAKLAELVAAAIAD 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLIAKLAELVAAAIAD 180

Qy  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSFGGLPSLAQVHAA 300
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSFGGLPSLAQVHAA 300

Qy  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360

Qy  361 SEGAAAGTEDAERAPVEADAGGGKVLVRNVV 392
        |||||||||||||||||||||||||||||||
Db  361 SEGAAAGTEDAERAPVEADAGGGKVLVRNVV 392
```

RESULT 2
US-09-134-001C-3461
; Sequence 3461, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14
NUMBER OF SEQ ID NOS: 5674
SEQ ID NO 3461
; LENGTH: 269
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3461

Query Match   7.5%; Score 148.5; DB 4; Length 269;
Best Local Similarity 34.0%; Pred. No. 6.8e-06;
Matches 48; Conservative 18; Mismatches 66; Indels 9; Gaps 4;

```
Qy  234 AGVPGLTG--ATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGF--- 288
            |||:||    ::|||  :|||||
Db   20 AGVPGSTGLPGSTGLPDSAGLPGSAGVPGSAGLPGSAGVPGSAGLPGS 79

Qy  289 GGLPSLAQV--HAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMH 346
        |||:|||    :|
Db   80 AGVPGSAGVPGSAGLPGSAGLPGSAG--VPGSAGLPGSAGVPGSAGLP 137

Qy  347 PSSGASKGTTTKKYSEGAAAG 367
        :::        :::
Db  138 GSAGLPGSAFVGPYSTISSTG 158
```

RESULT 3
US-09-060-756-727
; Sequence 727, Application US/09060756
; Patent No. 6183357
; GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen

APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/060,756
CURRENT FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 727
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Mycobacterium sp.
US-09-060-756-727

Query Match   6.9%; Score 135; DB 3; Length 318;
Best Local Similarity 27.5%; Pred. No. 0.00016;
Matches 58; Conservative 11; Mismatches 104; Indels 38; Gaps 7;

```
Qy  191 GTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVT 250
            ||  ||                                            :| ||
Db   14 GTPGAVGSGATGAPG--------GWLLGDGGAGGSGAAG--SGAPGGAGGAAGLWGTG 61

Qy  251 GLFGAAGLSASSG----------------LAHADSLASSASLPALAGIGGGSFGGLPSLAQVH 298
        ||                          |||||||:||:
Db   62 GAGGAGGSSAGGGAGGAGGAGGAGGWLLGDGGAGGIGGASTVLGGTGGGGVVGLWGAGGAG 121

Qy  299 AASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQG---------MGGPVGMGGMHPSSGAS 352
                                        :|:||          |:||
Db  122 GAGGTGLV--GGDGGAGGAGGTGGLLAGLIGAGGHGGTGLSTNGDGGVGG----AGGN 175

Qy  353 KGTTTKKYSEGAAAGTEDAERAPVEADAGGG 383
                            ||:|
Db  176 AGMLAGPGGAGGAGG--DGENLDTGDGDGAG 204
```

RESULT 4
US-09-670-314-727
; Sequence 727, Application US/09670314
; Patent No. 6492506
; GENERAL INFORMATION:
APPLICANT: Cole, Stewart
APPLICANT: Buchrieser-Brosch, Roland
APPLICANT: Gordon, Stephen
APPLICANT: Billault, Alain
TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
FILE REFERENCE: 3495-0169
CURRENT APPLICATION NUMBER: US/09/670,314
CURRENT FILING DATE: 2001-01-12
PRIOR APPLICATION NUMBER: 09/060,756
PRIOR FILING DATE: 1998-04-16
NUMBER OF SEQ ID NOS: 743
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 727
; LENGTH: 318
; TYPE: PRT
; ORGANISM: Mycobacterium sp.
US-09-670-314-727

Query Match   6.9%; Score 135; DB 4; Length 318;
Best Local Similarity 27.5%; Pred. No. 0.00016;
Matches 58; Conservative 11; Mismatches 104; Indels 38; Gaps 7;

```
Qy  191 GTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVT 250
            ||  ||                                            :| ||
Db   14 GTPGAVGSGATGAPG--------GWLLGDGGAGGSGAAG--SGAPGGAGGAAGLWGTG 61

Qy  251 GLFGAAGLSASSG----------------LAHADSLASSASLPALAGIGGGSFGGLPSLAQVH 298
        ||                          |||||||:||:
Db   62 GAGGAGGSSAGGGAGGAGGAGGAGGWLLGDGGAGGIGGASTVLGGTGGGGVVGLWGAGGAG 121
```

```
Qy  299  AASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQG------MGGPVGMGGWHPSSGAS  352
              |||    :  |:||                           :|:||
Db  122  GAGGTGLV--GGDGGAGGAGGTGGLLAGLIGAGGGHGGTGGLSTNGDGGVGG---AGGN  175

Qy  353  KGTTTKKYSEGAAAGTEDAERAPVEADAGGG  383
Db  176  AGMLAGPGGAGGAGG--DGENLDTGGDGGAG  204
```

RESULT 5
US-09-340-736E-1
; Sequence 1, Application US/09340736E
; Patent No. 6489946
; GENERAL INFORMATION:
; APPLICANT: ROTHSTEIN, ASER
; APPLICANT: KEELEY, FRED
; APPLICANT: ROTHSTEIN, STEVEN
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN ELASTIN
; TITLE OF INVENTION: AND OTHER FIBROUS PROTEINS
; FILE REFERENCE: 041082/0110
; CURRENT APPLICATION NUMBER: US/09/340,736E
; CURRENT FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 08/91,364
; PRIOR FILING DATE: 1997-08-07
; PRIOR APPLICATION NUMBER: 60/023,552
; PRIOR FILING DATE: 1996-08-07
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 731
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-340-736E-1

```
Query Match     6.8%;  Score 133.5;  DB 4;  Length 731;
Best Local Similarity  20.4%;  Pred. No. 0.00074;
Matches   86;  Conservative  35;  Mismatches 136;  Indels 165;  Gaps  15;

Qy   22  GIGIPN--QGGILYVSSLEYPEKALEELAAAFPPGDGWLGSAADXYAGKNRNHVNFFQELAD  79
             |||                                                        :
Db  117  GVGLPGVYPGGVL---------------------PGARFPPGVGVLPGV-------------  143

Qy   80  LDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPV-----AVDLTYIPVVGHALSAAFQ  134
             |:                                  |:||
Db  144  -------------PTGAGVKPKAPGVGGAFAGIPGVGPFGGPGQPGVPLGY-PIKAPKLPGGYG  192

Qy  135  APFCAGAM-------AVVGGA-LAYLVVKTLINATQLLKLLAKLAELVAAAIADIISDVA  186
                :                 |:                                       :
Db  193  LPYTTGKLPYGYPGPGVVAGAGKAGYPTGTGVGPQAAAAAAAAKFGAGAGVLPGVG  252

Qy  187  DIIKGTLGEVWEFITNALNGLKELMDKLITGWTGLFSRGWSNLESFFAGVPGLTGATSGL  246
                                                                 |||:| :
Db  253  G---------------------------------------------AGVPGVPGA-----  262

Qy  247  SQVTGLFGAAGLSASSGLAHADSLASSASLPALAG-IGGGSGFG--------GLPSL--  294
              |:||                                |||:|||       |:|:
Db  263  --IPGIGGIAGVGTPAAAAAAAAAKYGAAAGLVPGGPGFGPGVVGVPGACVPGVGV  320

Qy  295  --------AQVHAASTRQALRPRADGPVGAAAEQVG-----------------------  322
                    |                                                  :
Db  321  PGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAKYGARPGVGVGGIPTYGVGAGGFPG  380

Qy  323  ---CQSQLVSAQGSQGCMGGPGVGMG----GMHPSSGASKGTTTKKYSEG--AAAGTEDAE  372
                :|:||                  :|||:                 |:||
Db  381  FGVGVGGIPGVAGVPGVGVGPGVGVPGVGISPEAQAAAAAKAKYGVGTPAAAAAAKAA  440

Qy  373  RA  374
         :
Db  441  KA  442
```

RESULT 6
US-07-972-032-82

; Sequence 82, Application US/07972032
; Patent No. 5496712
; GENERAL INFORMATION:
; APPLICANT: Cappello, Joseph.
; APPLICANT: Ferrari, Franco A.
; TITLE OF INVENTION: HIGH MOLECULAR WEIGHT COLLAGEN-LIKE
; TITLE OF INVENTION: PROTEIN POLYMERS
; NUMBER OF SEQUENCES: 85
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bertram I. Rowland
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: CA 94111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/972,032
; FILING DATE: 19921105
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/791,960
; FILING DATE: 12-NOV-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Rowland, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A-55556-1/BIR;PROP-08-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; INFORMATION FOR SEQ ID NO: 82:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1077 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-972-032-82

```
Query Match     6.8%;  Score 133;  DB 1;  Length 1077;
Best Local Similarity  31.9%;  Pred. No. 0.0015;
Matches   51;  Conservative  10;  Mismatches  85;  Indels  14;  Gaps   5;

Qy  235  GVFGLTGA------TSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFG  289
Db   43  GAPGLQGAPGSQGAPGLQGAPGSQGAPGLQGAPGSQGAPGLQGAPGSQGAPGLQGAPGSQ  102

Qy  290  GLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQSQGMGGPVGMGGMHPSS  349
Db  103  GAPGLQGAPGSQG-APGLQGAPGSQG--GAPGLQGAPGSQGAPGLQGA-PGA  158

Qy  350  GASKGTTTKKYSEGA------AAGTEDAERAPVEADAGGGQ  384
Db  159  PGSQGAPGLQGAPGAPGSQGAPGLQGAPGSQGAPGLQGAPGLQ  198
```

RESULT 7
US-08-642-255-95
; Sequence 95, Application US/08642255
; Patent No. 5773349
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph.
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOHBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400

CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/642,255
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: ROWLAND, Bertram I.
REGISTRATION NUMBER: 20,015
REFERENCE/DOCKET NUMBER: A55556-3/BIR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-8700
TELEFAX: (415) 494-8771
TELEX: 910 277299 FHT UR
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 1077 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-642-255-95

Query Match   6.8%; Score 133; DB 1; Length 1077;
Best Local Similarity 31.9%; Pred. No. 0.0015;
Matches   51; Conservative   10; Mismatches   85; Indels   14; Gaps   5;

Qy   235 GVPGLTGA----------TSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFG 289
Db   43 GAPGLQGAPGAPGSQGAPGLQGAPGAPGSQGAPGLQGAPGAPGSQGAPGLQGAPGAPGSQ 102

Qy   290 GLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSS 349
Db   103 GAPGLQGAPGAPGSQG-APGLQGAPGAPGSQ--GAPGLQGAPGSQGAPGLQGA-PGA 158

Qy   350 GASKGTTTKKYSEGA-----AAGTEDAERAPVEADAGGGQ 384
Db   159 PGSQGAPGLQGAPGAPGSQGAPGAPGSQGAPGLQ 198

RESULT 8
US-08-911-364-1
; Sequence 1, Application US/08911364
; Patent No. 5959106
GENERAL INFORMATION:
APPLICANT: ROTHSTEIN, Aser
APPLICANT: KEELY, Fred W.
APPLICANT: ROTHSTEIN, Steven J.
TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN
TITLE OF INVENTION: ELASTIN AND OTHER FIBROUS PROTEINS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: FOLEY & LARDNER
STREET: 3000 K Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/911,364
FILING DATE: 07-AUG-1997

CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/023,552
FILING DATE: 07-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: Bent, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 041082/0104
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 672-5300
TELEFAX: (202) 672-5399
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 731 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-911-364-1

Query Match   6.6%; Score 129.5; DB 2; Length 731;
Best Local Similarity 20.3%; Pred. No. 0.0018;
Matches   86; Conservative   36; Mismatches 134; Indels 167; Gaps   15;

Qy   22 GIGIPN--QGGILYSSLEYFEKALELAAAFPGDGWLGSAADKYAGKNRNHVNFFQELAD 79
Db   117 GVGLPGVPGGVL----------------PGARPGVGVLPGV------------------ 143

Qy   80 LDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPV----AVDLTYIPVVGHALSAAFQ 134
Db   144 ---------------PTGAGVKPKAPGVGGAFAGIPGVGPFGGPQPGVPLGY-FIKAPKLPGGYG 192

Qy   135 APFCAGAM------AVVGGA-LAYLVVKTLINATQLLKLLAKLAELVAAAIADISDVA 186
Db   193 LPYTTGKLPYGYGPGVGVAGRAGKAGYPTGTGVGPQAAAAAAAKAAKFGAGAAGVLPGVG 252

Qy   187 DIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGVPGLTGATSGL 246
Db   253 G---AGVPGVPGA---- 262

Qy   247 SQVTGLFGAAGLSASSGLAHADSLASSASLPALAG-IGGGSGFG------GLPSL-- 294
Db   263 --IPGIGGIAGVGTPAAAAAAAAKAAKYGAAAGLVPGGPGFGPGVGVPGVGVGV 320

Qy   295 --------AQVHAASTRQALRPRADGPVGAAAEQVG---------------------- 322
Db   321 PGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARPGVGVGGIPTYGVGAGGFPG 380

Qy   323 --------GQSQLVSAQGSQQMGGPVGMGGMHPSSGASKGTTTKKYSEG--AAAGTEDA 371
Db   381 FGVGVGIPGVAGVPSVGGVPGVGGVPGV-GISPEAQAAAAAKAAKYGVGTPAAAAAKAA 439

Qy   372 ERA 374
Db   440 AKA 442

RESULT 9
US-08-464-700-2
; Sequence 2, Application US/08464700
; Patent No. 6232458
GENERAL INFORMATION:
APPLICANT: WEISS, ANTHONY S
APPLICANT: MARTIN, STEPHEN L
TITLE OF INVENTION: SYNTHETIC POLYNUCLEOTIDES
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: Howson and Howson
STREET: Spring House Corporate Cntr, PO Box 457
CITY: Spring House
STATE: Pennsylvania
COUNTRY: USA
ZIP: 19477

```
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/464,700
;     FILING DATE: 7-JUN-1995
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: AU PL6520
;     FILING DATE: 22-DEC-1992
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: AU PL9661
;     FILING DATE: 28-JUN-1993
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: PCT/AU93/06655
;     FILING DATE: 16-DEC-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Bak, Mary E.
;     REGISTRATION NUMBER: 31,215
;     REFERENCE/DOCKET NUMBER: GHC3USA
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 215-540-9200
;     TELEFAX: 215-540-5818
;   INFORMATION FOR SEQ ID NO: 2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 733 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
;       MOLECULE TYPE: protein
; US-08-464-700-2

Query Match        6.6%;   Score 129.5; DB 3; Length 733;
Best Local Similarity  20.3%;  Pred. No. 0.0018;
Matches  86;  Conservative  36;  Mismatches 134;  Indels 167;  Gaps 15;

Qy    22 GIGIPN--QGGILYSSLEYFEKALELAAAFPGDGWLGSAADKYAGKNRNHVNFFQELAD  79
              |::| |                              ||||
Db   119 GVGLPGVYPGGVL---------------PGARFPGGVVLPGV------------------ 145

Qy    80 LDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPV----AVDLTYIPVVGHALSAAFQ 134
                                                     |:|    :
Db   146 ------------PTGAGVPKPAPGVGGAFAGIPGVGPGQPGVPLGY-PIKAPKLPGGYG 194

Qy   135 APFCAGAM------AVVGGA-LAYLVVKTLINATQLLKLLAELVAAAIADIISDVA     186
         |:                 :::|                    | |::
Db   195 LPYTTGKLFYGYGPGGVAGAAGKAGYPTGTGVGPQAAAAAAAKFGAGAAGVLPGVG     254

Qy   187 DIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGL 246
                                                              ||||||
Db   255 G------------------------------------------AGVPGVPGA------   264

Qy   247 SQVTGLFGAGLSASSGLAHADSLASSASLPALAG-IGGGSGFG------------GLPSL-- 294
             : :                 ::::                            :|:
Db   265 --IPGIGGIAGVGTPAAAAAAAAKAAKYGAAAGLVPGPGFGPGVVVPGAGVPGVGV      322

Qy   295 ------AQVHAASTRQALRPRADGPVGAAAEQVG                            322
Db   323 PGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAKYGARPGVGVGGIPTYGVGAGGFPG      382

Qy   323 ------GQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKYSEG--AAAGTEDA     371
              | |: |       :  :  ::     |  :  |   ::    |
Db   383 FGVGVGGIPGVAGVPSVGGVPGVGGVPGV-GISPEAQAAAAAKAAKYGVGTPAAAAAKAA  441

Qy   372 ERA 374
             :
Db   442 AKA 444

RESULT 10
US-08-095-734-2
; Sequence 2, Application US/08095734
```

```
; Patent No. 5807723
; GENERAL INFORMATION:
;   APPLICANT: Aldovini, Anna
;   APPLICANT: Young, Richard A.
;   TITLE OF INVENTION: Homologously Recombinant Slow Growing
;   TITLE OF INVENTION: Mycobacteria and Uses Therefor
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;     STREET: Two Militia Drive
;     CITY: Lexington
;     STATE: MA
;     COUNTRY: USA
;     ZIP: 02173
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/095,734
;     FILING DATE: 22-JUL-1993
;     CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Granahan, Patricia
;     REGISTRATION NUMBER: 32,227
;     REFERENCE/DOCKET NUMBER: WHI93-11M
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-861-6240
;     TELEFAX: 617-861-9540
;   INFORMATION FOR SEQ ID NO: 2:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 1271 amino acids
;       TYPE: amino acid
;       STRANDEDNESS: single
;       TOPOLOGY: linear
;       MOLECULE TYPE: protein
; US-08-095-734-2

Query Match        6.6%;   Score 129.5; DB 1; Length 1271;
Best Local Similarity  21.4%;  Pred. No. 0.0039;
Matches  96;  Conservative  52;  Mismatches 155;  Indels 145;  Gaps 19;

Qy    22 GIGIPNQGGILYS--------SLEYFEKA-LEELAAAFPGDGWLGSAADKYAGKNRNHVN   72
Db   831 GAGV--GPMLISATQNGELSAQYAEAASEVEELLGVVASEGWGQQAEEALVAAYMPFLA   887

Qy    73 FF---QELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLTYIPVVGHA  128
Db   888 WLIQASADCVEMAAQQHAVIEAYTAAVELMPTQVELAANQIKLAVLVATNFFGNTIPIA  947

Qy   129 LSAAFQAPPCAGAMAVVGGALAYLIVVKTLINA---TQLLKLLAKLIAELVAAAIADIISD-  184
Db   948 INEAEYVEMVVRAATM--ATYSTVSRSALSAVMPHTSPPPLLILKSDELLPDTGEDSDEDG 1005

Qy   185 -------------VADIIKG-TLGE-VWEFITNALNGL------KELWDKLT----      215
Db  1006 HNHGGHSGHSGGHARMIDNFFAEILRGVSAGRIVWDPVNGTLNGLDYDDVYVPGHAIW--- 1061

Qy   216 GWVT-GL--FSRGWSNLESFFAGVPG------                            238
Db  1062 -WLARGLEFFQDGEQFGELLFTNPTGAFQFLLYVVVVDLPTHIAQIATWLGQYPQLLSAA 1120

Qy   239 LTGATSGLSQVTGLFGAGLSASSGLA-----HADSLASSASLPALAGIG----       283
Db  1121 LTGVIAHLGAITGLAGLSGLSAIPSAAIPAVVPELTPVAAAPPMLAVAGVGPAVAAAEQVG 1180

Qy   284 ------------GGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVG-         322
Db  1181 PASAPAPAAAGATAAGPTPPATGFGGLPAL--------PGRRWRPRNRVRLGTVGPRQGR  1233

Qy   323 GQSQLVSAQGSQGMG-------GPVG 341
```

```
Db   831 GAGV--...GPMLISATQNGELSAQYABAASEVEELLGVVASEGWQGQAAEAIVAAYMPFLA 887
          | |:   | |:| |          |||    :| :|  |          :|
Qy    73 FF----QELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLTYIPVVGHA 128

Db   888 WLIQASADCVEMAAQQHAVIEAYTAAVELMPTQVELAANQIKLAVLVATNFFGINTIPIA 947
          :                           |:  |:   |  :: :  :
Qy   129 LSAAFQAPFCAGAMAVVGGALAYLVKTLINA--TQLLKLLAKLAELVAAAIADIISD- 184

Db   948 INEAEYVEMWVRAATM--ATYSTVSRSALSAMPHTSPPPILKSDELLPDTGEDSDEDG 1005
          :|:   |              |||  | :|
Qy   185 ------VADIKG-TLGE-VWEFITNALNGL------KELWDKLT 215

Db  1006 HNHGGHSHGGHARMIDNFPFAEILRGVSAGRIVWDPVNGTLNGLDYDDYVVPGHAIW---- 1061

Qy   216 -WLARGLEFFQDGEQFGELLFTNPTGAFQFLLYVVVDLPTHIAQIATWLGQYPQLLSAA 238

Db  1062 -WLARGLEFFQDGEQFGELLFTNPTGAFQFLLYVVVDLPTHIAQIATWLGQYPQLLSAA 1120

Qy   239 LTGATSGLSQVTGLFGAAGLGASSGLA-----HADSLASSASLPALAGIG------ 283
          ||| :|  :| |||| |||  |:|  ||
Db  1121 LTGVIAHLGATTGLAGLSGLSAIPSAAIPAVVPELTPVAAAPPMLAVAGVGPAVAAPGML 1180

Qy   284 ---...GGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVG- 322
                  |||:||||| :|         |||:|
Db  1181 PASAPAPAAAGATAAGPTPPATGFCGLPAL------PGRWRPRNRVRLGTVGPRQGR 1233

Qy   323 GQSQLVSAQGSQGMG------GPVG 341
          |:|:|  |||  :|        ||||
Db  1234 GVRFRCSRVGGPGLGACAGACCTAGPLG 1261
```

RESULT 11
US-08-444-623-2
; Sequence 2, Application US/08444623
; Patent No. 5866403
GENERAL INFORMATION:
- APPLICANT: Aldovini, Anna
- APPLICANT: Young, Richard A.
- TITLE OF INVENTION: Homologously Recombinant Slow Growing
- TITLE OF INVENTION: Mycobacteria and Uses Therefor
- NUMBER OF SEQUENCES: 2
- CORRESPONDENCE ADDRESS:
- ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
- STREET: Two Militia Drive
- CITY: Lexington
- STATE: MA
- COUNTRY: USA
- ZIP: 02173
- COMPUTER READABLE FORM:
- MEDIUM TYPE: Floppy disk
- COMPUTER: IBM PC compatible
- OPERATING SYSTEM: PC-DOS/MS-DOS
- SOFTWARE: PatentIn Release #1.0, Version #1.25
- CURRENT APPLICATION DATA:
- APPLICATION NUMBER: US/08/444,623
- FILING DATE:
- CLASSIFICATION: 435
- PRIOR APPLICATION DATA:
- APPLICATION NUMBER: 08/095,734
- FILING DATE: 22-JUL-1993
- APPLICATION NUMBER: 07/711,334
- FILING DATE: 06-JUN-1991
- APPLICATION NUMBER: 07/367,894
- FILING DATE: 19-JUN-1989
- APPLICATION NUMBER: PCT/US90/03451
- FILING DATE: 18-JUN-1990
- APPLICATION NUMBER: PCT/US89/02962
- FILING DATE: 07-JUL-1989
- APPLICATION NUMBER: 07/361,944
- FILING DATE: 05-JUN-1989
- APPLICATION NUMBER: 07/223,089
- FILING DATE: 22-JULY-1988
- APPLICATION NUMBER: 07/216,390
- FILING DATE: 07-JUL-1988
- APPLICATION NUMBER: 07/163,546
- FILING DATE: 03-MAR-1988
- APPLICATION NUMBER: PCT/US88/00614
- FILING DATE: 29-FEB-1988
- APPLICATION NUMBER: 07/020,451
- FILING DATE: 02-MAR-1987
- ATTORNEY/AGENT INFORMATION:
- NAME: Granahan, Patricia
- REGISTRATION NUMBER: 32,227
- REFERENCE/DOCKET NUMBER: WHI93-11MZ
- TELECOMMUNICATION INFORMATION:
- TELEPHONE: 617-861-6240
- TELEFAX: 617-861-9540
- INFORMATION FOR SEQ ID NO: 2:
- SEQUENCE CHARACTERISTICS:
- LENGTH: 1271 amino acids
- TYPE: amino acid
- STRANDEDNESS: single
- TOPOLOGY: linear
- MOLECULE TYPE: protein
US-08-444-623-2

Query Match   6.6%; Score 129.5; DB 2; Length 1271;
Best Local Similarity 21.4%; Pred. No. 0.0039;
Matches 96; Conservative 52; Mismatches 155; Indels 145; Gaps 19;

Qy   22 GIGIPNQGGILYS------SLEYFEKA--LEELAAAFPGDGWLGSAADKYAGKNRNHVN 72

RESULT 12
US-08-471-869-2
; Sequence 2, Application US/08471869
; Patent No. 6022745
GENERAL INFORMATION:
- APPLICANT: Aldovini, Anna
- APPLICANT: Young, Richard A.
- TITLE OF INVENTION: Homologously Recombinant Slow Growing
- TITLE OF INVENTION: Mycobacteria and Uses Therefor
- NUMBER OF SEQUENCES: 2
- CORRESPONDENCE ADDRESS:
- ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
- STREET: Two Militia Drive
- CITY: Lexington
- STATE: MA
- COUNTRY: USA
- ZIP: 02173
- COMPUTER READABLE FORM:
- MEDIUM TYPE: Floppy disk
- COMPUTER: IBM PC compatible
- OPERATING SYSTEM: PC-DOS/MS-DOS
- SOFTWARE: PatentIn Release #1.0, Version #1.30
- CURRENT APPLICATION DATA:
- APPLICATION NUMBER: US/08/471,869
- FILING DATE: 07-JUN-1995
- CLASSIFICATION: 435
- PRIOR APPLICATION DATA:
- APPLICATION NUMBER: PCT/US94/08267
- FILING DATE: 22-JUL-1994
- PRIOR APPLICATION DATA:
- APPLICATION NUMBER: US 08/095,734
- FILING DATE: 22-JUL-1993
- PRIOR APPLICATION DATA:
- APPLICATION NUMBER: US 07/711,334
- FILING DATE: 06-JUN-1991
- PRIOR APPLICATION DATA:
- APPLICATION NUMBER: US 07/367,894
- FILING DATE: 19-JUN-1989
- PRIOR APPLICATION DATA:
- APPLICATION NUMBER: PCT/US90/03451
- FILING DATE: 18-JUN-1990

PRIOR APPLICATION DATA:
  APPLICATION NUMBER: PCT/US89/02962
  FILING DATE: 07-JUL-1989
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/361,944
  FILING DATE: 05-JUN-1989
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/223,089
  FILING DATE: 22-JUL-1988
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/216,390
  FILING DATE: 07-JUL-1988
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/163,546
  FILING DATE: 03-MAR-1988
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: PCT/US88/00614
  FILING DATE: 29-FEB-1988
PRIOR APPLICATION DATA:
  APPLICATION NUMBER: US 07/020,451
  FILING DATE: 02-MAR-1987
ATTORNEY/AGENT INFORMATION:
  NAME: Granahan, Patricia
  REGISTRATION NUMBER: 32,227
  REFERENCE/DOCKET NUMBER: WHI93-11MA2
TELECOMMUNICATION INFORMATION:
  TELEPHONE: 617-861-6240
  TELEFAX: 617-861-9540
INFORMATION FOR SEQ ID NO: 2:
  SEQUENCE CHARACTERISTICS:
    LENGTH: 1271 amino acids
    TYPE: amino acid
    STRANDEDNESS: single
    TOPOLOGY: linear
  MOLECULE TYPE: protein
US-08-471-869-2

Query Match   6.6%; Score 129.5; DB 3; Length 1271;
Best Local Similarity 21.4%; Pred. No. 0.0039;
Matches 96; Conservative 52; Mismatches 155; Indels 145; Gaps 19;

```
Qy    22 GIGIPNQGGILYS------SLEYFEKA--LEELAAAFPGDGWLGSAADKYAGKNRNHVN   72
         |:|   :||||:        ::|::      :|:||  ::|
Db   831 GAGV---GPMLISATQNGELSAQYAEAASEVEELLGVVASEGWQGQAAEALVAAYMPFLA  887

Qy    73 FF----QELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLTYIPVVGHA  128
         ||     |:|  :  :||||||||||||  ||:|      |::|
Db   888 WLIQASADCVEMAAQQHAVIEAYTAAVELMPTQVELAANQIKLAVLVATNFFGINTIPIA  947

Qy   129 LSAAFQAPFCAGAMAVVGGALAYLVVKTLINA--TQLLKLLAKLAELVAAAIADIISD-  184
         :  :       : ::|||:|:|:||  :|||::  :|:|
Db   948 INEAEYVEMVVRAATTM--ATYSTVSRSALSAMPHTSPPPLILKSDELLPDTGEDSDEDG 1005

Qy   185 --------------VADIIKG-TLGE-VWEFITNALNGL-------KELWDKLT        215
                       :|:|||||  ||:|     ||||
Db  1006 HNHGGHSGGHARMIDNFFAEILRGVSAGRIVWDPVNGTLNGLDYDDYVYPGHAIW---- 1061

Qy   216 GWVT-GL--FSRGWSNLESFFAGVPG-----------------------         238
         :||  |:    :|
Db  1062 -WLARGLEFFQDGEQFGELLFTNPTGAFQFLLYVVVVDLPTHIAQIATWLGQYPQLLSAA 1120

Qy   239 LTGATSGLSQVTGLFGAAGLSASSGLA-----HADSLASSASLPALAGIG---------  283
         ||| :  :|||:|||||||:|::::||     :||:|||::|:|||||:|
Db  1121 LTGVIAHLGAITGLAGLSGLSAIPSAAIPAVVPELTPVAAAPPMLAVAGVGPAVAAPGML 1180

Qy   284 ----------GGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVG-        322
                   |||||||||::|:||  ::|:|::|:
Db  1181 PASAPAPAAAAGATAAGPTPPATGFGGLPAL------PGRRWRPRNRVRLGTVGPRQGR 1233

Qy   323 GQSQLVSAQGSQGMG--------GPVG                                 341
         |:|:|| | |:|:||        ||||
Db  1234 GVRFRCSRVGGPGLGACAGACCTAGPLG                               1261
```

GENERAL INFORMATION:
```
; APPLICANT:
;   TITLE OF INVENTION: Homologously Recombinant Slow Growing
;   TITLE OF INVENTION: Mycobacteria and Uses Therefor
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
;     STREET: Two Militia Drive
;     CITY: Lexington
;     STATE: MA
;     ZIP: 02173
;     COUNTRY: USA
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: PCT/US94/08267
;     FILING DATE:
;     CLASSIFICATION:
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/095,734
;     FILING DATE: 22-JUL-1993
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Granahan, Patricia
;     REGISTRATION NUMBER: 32,227
;     REFERENCE/DOCKET NUMBER: WHI93-11MA PCT
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 617-861-6240
;     TELEFAX: 617-861-9540
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 1271 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
PCT-US94-08267-2

Query Match
Best Local Similarity  21.4%;  Score 129.5;  DB 5;  Length 1271;
  Matches  96;  Conservative  52;  Mismatches 155;  Indels 145;  Gaps  19;

Qy   22 GIGIPNQGGILYS------SLEYFEKA--LBELAAAFPGDGWLGSAADKYAGKNRRHVN  72
Db  831 GAGV--GEMLISATONGELSAQYAEAASEVEBLLGVVASEGWQGQAAEALVAAYMPFLA 887

Qy   73 FF---QELADLDRQLISLIHDQANAVOTTRDILEGAKKGLEFVRPVAVDLTYIPVVGHA 128
Db  888 WLIQASADCVEMAAQQHAVIEAYTAAVELMPTQVELAANQIKLAVLVATNFFGINTIPIA 947

Qy  129 LSAAFQAPFCAGAMAVVGGALAYLVVKTLINA---TQLLKLLAKLAELVARAIADIISD- 184
Db  948 INEAEYEVEMVRAATTM-ATYSTVSRSALSAMPHTSPPLLIKSDELLPDTGEDSDEDG 1005

Qy  185 ------------VADIKG-TLGE-VWEFITNALNGL-------------KELWDKLT 215
Db 1006 HNHGGHSHGGHARMIDNFFAEILRGVSAGRIVWDPVNGTLNGLDYDDYVYPGHAIW---- 1061

Qy  216 GWVT-GL--FSRGWSNLESFFAGVPG------------------------GPVG 238
Db 1062 -WLARGLEFFQDEGEQFGELLFTNPTGAFQFLLYVVVVDLPTHIAQIATWLGQYPQLLSAA 1120

Qy  239 LTGATSGLSQVTGLFGAAGLSASSGLA-----HADSLASSASLPALAGIG------- 283
Db 1121 LTGVIAHLGAITGLAGLSGLSGLSAIPSAAIPAVVPELTPVAAAPPMLAVAGVGPAVAAPGML 1180

Qy  284 ----GGSGFGGLPSLAQVHAASTRQALRPRADGPVGAL-----PGRRWRPRNRVRLGTVGPRQGR 322
Db 1181 PASAPAPAAAAGAGATAAGPTPPATGFGGLPAL----- 1233

Qy  323 GQSQLVSAQGSQGMG--------GPVG 341
Db 1234 GVRFRCSRVGPGLGACAGACCTAGPLG 1261
```

```
US-08-209-747-2

Query Match
Best Local Similarity  24.6%;  Score 128.5;  DB 1;  Length 832;
  Matches  70;  Conservative  28;  Mismatches 129;  Indels  57;  Gaps  11;

Qy  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
Db   11 YVSTLGNAISDA---------SAYANALSAIGNVLANS-------GSISESTASSAA-  52

Qy  181 IISDVADIIKGTL---GEVWEFITNALNGLKELWDKLTGWVTG---LFSRGWSNLESFFA 234
Db   53 --SSAASVTTLTSYGPAVFYAPSASSG------GYGAGAVAAAGAAGAGGYGR 101

Qy  235 GVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASSLPALAGIG---GGSGFGGL 291
Db  102 GAGGY-GGQGGYGAGAGAAAAGAGAGGYGRGAGAGAGAAAGAGAAAGAGAGGYGGQ 160

Qy  292 PSLAQVHAASTRQALRPRADGPVG-----AAAQVGGQSQLVSAQGSQGMGGPV 340
Db  161 GGYGAGAGAAAAGAGAGAGAAAGAGAGGYGRGAGAGAGAGAAAGAGAAAGAGAGGYG-----GQGGYGAGA 214
```

Thu Nov 6 11:33:26 2003

```
Qy    341  GMGGMHPSSGASKGTTTKKYSEGAAAGTEDAERAPVEADAGGGQ 384
           |||:||||  ||||||||  |||||||||  |:||||||||
Db    215  GAGAAAAAGAGSG-GAGGYGRGAGAGAGAAAGAGAGAGSYGGQ 257
```

Search completed: November 5, 2003, 17:01:31
Job time : 24 secs

This Page Blank (uspto)

```
              GenCore version 5.1.6

          Copyright (c) 1993 - 2003 Compugen Ltd.


OM protein - protein search, using sw model


Run on:      November  5, 2003, 17:02:43 ; Search time 44 Seconds
                                  (without alignments)
                                  1530.130 Million cell updates/sec

Title:         US-09-688-672A-14
Perfect score: 1969
Sequence:      1 MSRAFIIDPTISAIDGLYDL........RAPVEADAGGGQKVLVRNVV 392

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      644079 seqs, 171749292 residues

Total number of hits satisfying chosen parameters:      644079

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     Published_Applications_AA:*
                1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
                2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
                3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
                4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
                5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
                6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
                7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
                8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
                9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
               10: /cgn2_6/ptodata/2/pubpaa/US05B_PUBCOMB.pep:*
               11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
               12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
               13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
               14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
               15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
               16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
               17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
               18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1969	100.0	392	9	US-09-073-009-138	Sequence 138, App
2	1969	100.0	392	9	US-09-793-306-138	Sequence 138, App
3	1969	100.0	392	12	US-10-098-732A-29	Sequence 29, Appl
4	1969	100.0	392	16	US-10-080-170-610	Sequence 610, App
5	1258	63.9	394	16	US-10-080-170-53	Sequence 53, Appl
6	140.5	7.1	256	11	US-09-820-843A-18	Sequence 18, Appl
7	135	6.9	318	12	US-10-259-678-727	Sequence 727, App
8	133.5	6.8	731	12	US-09-964-662-1	Sequence 1, Appl
9	131.5	6.7	446	15	US-10-155-76-13993	Sequence 13993, A
10	130	6.6	1079	11	US-09-820-843A-20	Sequence 20, Appl
11	129.5	6.6	730	11	US-09-961-403-8	Sequence 8, Appli
12	129	6.6	1149	12	US-10-097-111-283	Sequence 283, App
13	127	6.4	585	10	US-09-738-626-6032	Sequence 6032, Ap
14	126.5	6.4	334	12	US-10-259-678-728	Sequence 728, App
15	126	6.4	749	15	US-10-156-761-11729	Sequence 11729, A
16	123.5	6.3	651	9	US-09-861-597-1	Sequence 1, Appli
17	123.5	6.3	651	12	US-10-414-760-1	Sequence 1, Appli
18	121.5	6.2	1488	10	US-09-738-626-5495	Sequence 5495, Ap
19	120.5	6.1	1321	15	US-10-184-644-371	Sequence 371, App
20	120.5	6.1	1321	15	US-10-184-634-371	Sequence 371, App
21	120	6.1	766	15	US-10-096-986-88	Sequence 88, Appl
22	120	6.1	979	15	US-10-096-986-89	Sequence 89, Appl
23	120	6.1	1894	12	US-10-199-672-97	Sequence 97, Appl
24	120	6.1	1894	12	US-10-187-749-97	Sequence 97, Appl
25	120	6.1	1894	12	US-10-194-457-97	Sequence 97, Appl
26	120	6.1	1894	12	US-10-184-642-97	Sequence 97, Appl
27	120	6.1	1894	12	US-10-196-747-97	Sequence 97, Appl
28	120	6.1	1894	12	US-10-173-689-97	Sequence 97, Appl
29	120	6.1	1894	12	US-10-173-690-97	Sequence 97, Appl
30	120	6.1	1894	12	US-10-173-691-97	Sequence 97, Appl
31	120	6.1	1894	12	US-10-173-692-97	Sequence 97, Appl
32	120	6.1	1894	12	US-10-173-694-97	Sequence 97, Appl
33	120	6.1	1894	12	US-10-173-699-97	Sequence 97, Appl
34	120	6.1	1894	12	US-10-173-698-97	Sequence 97, Appl
35	120	6.1	1894	12	US-10-173-707-97	Sequence 97, Appl
36	120	6.1	1894	12	US-10-174-569-97	Sequence 97, Appl
37	120	6.1	1894	12	US-10-174-583-97	Sequence 97, Appl
38	120	6.1	1894	12	US-10-174-587-97	Sequence 97, Appl
39	120	6.1	1894	12	US-10-174-589-97	Sequence 97, Appl
40	120	6.1	1894	12	US-10-174-591-97	Sequence 97, Appl
41	120	6.1	1894	12	US-10-175-736-97	Sequence 97, Appl
42	120	6.1	1894	12	US-10-175-742-97	Sequence 97, Appl
43	120	6.1	1894	12	US-10-175-744-97	Sequence 97, Appl
44	120	6.1	1894	12	US-10-175-745-97	Sequence 97, Appl
45	120	6.1	1894	12	US-10-175-748-97	Sequence 97, Appl

ALIGNMENTS

```
RESULT 1
US-09-073-009-138
; Sequence 138, Application US/09073009
; Patent No. US2001001288A1
, GENERAL INFORMATION:
,   APPLICANT: Alderson, Mark
,   APPLICANT: Dillon, Davin C.
,   APPLICANT: Skeiky, Yasir A.W.
,   APPLICANT: Campos-Neto, Antonio
,   TITLE OF INVENTION: COMPOUNDS AND DIAGNOSIS OF
,   NUMBER OF SEQUENCES: 144
,   CORRESPONDENCE ADDRESS:
,     ADDRESSEE: SEED and BERRY
,     STREET: 6300 Coumbia Center, 701 Fifth Ave.
,     CITY: Seattle
,     STATE: Washington
,     COUNTRY: US
,     ZIP: 98104
,   COMPUTER READABLE FORM:
,     MEDIUM TYPE: Floppy disk
,     COMPUTER: IBM PC compatible
,     OPERATING SYSTEM: PC-DOS/MS-DOS
,     SOFTWARE: PatentIn Release #1.0, Version #1.30
,   CURRENT APPLICATION DATA:
,     APPLICATION NUMBER: US/09/073,009
,     FILING DATE: 05-MAY-1998
,     CLASSIFICATION:
,   ATTORNEY/AGENT INFORMATION:
,     NAME: Maki, David J.
,     REGISTRATION NUMBER: 31,392
,     REFERENCE/DOCKET NUMBER: 210121.441C1
,   TELECOMMUNICATION INFORMATION:
,     TELEPHONE: 206-622-4900
,     TELEFAX: 206-682-6031
, INFORMATION FOR SEQ ID NO: 138:
,   SEQUENCE CHARACTERISTICS:
,     LENGTH: 392 amino acids
,     TYPE: amino acid
```

```
;  STRANDEDNESS: single
;  TOPOLOGY: linear
;  MOLECULE TYPE: protein
US-09-073-009-138

Query Match             100.0%;  Score 1969; DB 9; Length 392;
Best Local Similarity   100.0%;  Pred. No. 9.4e-166;
Matches  392;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120

Qy  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGVPGLT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGVPGLT 240

Qy  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300

Qy  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY 360

Qy  361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
        |||||||||||||||||||||||||||||||
Db  361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392


RESULT 2
US-09-793-306-138
;  Sequence 138, Application US/0979330G
;  Patent No. US2002009820OA1
;  GENERAL INFORMATION:
;   APPLICANT: Campos-Neto, Antonio
;   APPLICANT: Skeiky, Yasir
;   APPLICANT: Ovendale, Pamela
;   APPLICANT: Jen, Shyian
;   APPLICANT: Lodes, Michael
;   APPLICANT: Corixa Corporation
;   TITLE OF INVENTION: Compounds and Methods for Diagnosis and Immunotherapy
;   TITLE OF INVENTION: of Tuberculosis
;   FILE REFERENCE: 014058-0087400US
;   CURRENT APPLICATION NUMBER: US/09/793,306
;   CURRENT FILING DATE: 2001-02-26
;   PRIOR APPLICATION NUMBER: US 60/185,037
;   PRIOR FILING DATE: 2000-02-25
;   PRIOR APPLICATION NUMBER: US 60/223,828
;   PRIOR FILING DATE: 2000-08-08
;   NUMBER OF SEQ ID NOS: 164
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 138
;   LENGTH: 392
;   TYPE: PRT
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: hTcc#1
US-09-793-306-138

Query Match             100.0%;  Score 1969; DB 9; Length 392;
Best Local Similarity   100.0%;  Pred. No. 9.4e-166;
Matches  392;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120

Qy  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGVPGLT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGVPGLT 240

Qy  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300

Qy  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY 360
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY 360

Qy  361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
        |||||||||||||||||||||||||||||||
Db  361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392


RESULT 3
US-10-098-732A-29
;  Sequence 29, Application US/10098732A
;  Publication No. US2003017529A1
;  GENERAL INFORMATION:
;   APPLICANT: Skeiky, Yasir
;   APPLICANT: Brannon, Mark
;   APPLICANT: Guderian, Jeffrey
;   APPLICANT: Corixa Corporation
;   TITLE OF INVENTION: Heterologous Fusion Protein Constructs Comprising a
;   TITLE OF INVENTION: Leishmania Antigen
;   FILE REFERENCE: 014058-012010US
;   CURRENT APPLICATION NUMBER: US/10/098,732A
;   CURRENT FILING DATE: 2003-04-29
;   PRIOR APPLICATION NUMBER: US 60/275,837
;   PRIOR FILING DATE: 2001-03-13
;   NUMBER OF SEQ ID NOS: 80
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 29
;   LENGTH: 392
;   TYPE: PRT
;   ORGANISM: Mycobacterium tuberculosis
;   FEATURE:
;   OTHER INFORMATION: MTB40 (HTCC#1)
US-10-098-732A-29

Query Match             100.0%;  Score 1969; DB 12; Length 392;
Best Local Similarity   100.0%;  Pred. No. 9.4e-166;
Matches  392;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120

Qy  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGVPGLT 240
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGVPGLT 240
```

```
Qy  241  GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA  300

Qy  301  STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY  360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY  360

Qy  361  SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV  392
         |||||||||||||||||||||||||||||||
Db  361  SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV  392
```

RESULT 4
US-10-080-170-610
; Sequence 610, Application US20030129601A1
; Publication No. US20030129601A1
GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 610
; LENGTH: 392
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-10-080-170-610

```
Query Match        100.0%;  Score 1969;  DB 16;  Length 392;
Best Local Similarity 100.0%;  Pred. No. 9.4e-166;
Matches  392;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy    1  MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALELAAAFPGDGWLGSAA   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALELAAAFPGDGWLGSAA   60

Qy   61  DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT  120
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT  120

Qy  121  YIPVVGHALSAAFQAPFCAGAMAVVGGALAYIVVKTLINATQLLKLLAKLAELVAAAIAD  180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  YIPVVGHALSAAFQAPFCAGAMAVVGGALAYIVVKTLINATQLLKLLAKLAELVAAAIAD  180

Qy  181  IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181  IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT  240

Qy  241  GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA  300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA  300

Qy  301  STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY  360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASKGTTTKKY  360

Qy  361  SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV  392
         |||||||||||||||||||||||||||||||
Db  361  SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV  392
```

RESULT 5
US-10-080-170-53
; Sequence 53, Application US10080170
; Publication No. US20030129601A1

GENERAL INFORMATION:
; APPLICANT: COLE, S.T.
; TITLE OF INVENTION: COMPARATIVE MYCOBACTERIAL GENOMICS AS A TOOL FOR
; TITLE OF INVENTION: IDENTIFYING TARGETS FOR THE DIAGNOSIS, PROPHYLAXIS OR
; TITLE OF INVENTION: TREATMENT OF MYCOBACTERIOSES
; FILE REFERENCE: 03495.0218
; CURRENT APPLICATION NUMBER: US/10/080,170
; CURRENT FILING DATE: 2002-06-10
; PRIOR APPLICATION NUMBER: 60/270,123
; PRIOR FILING DATE: 2001-02-22
; NUMBER OF SEQ ID NOS: 652
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 53
; LENGTH: 394
; TYPE: PRT
; ORGANISM: Mycobacterium leprae
US-10-080-170-53

```
Query Match        63.9%;   Score 1258;  DB 16;  Length 394;
Best Local Similarity 62.7%;  Pred. No. 7e-103;
Matches  247;  Conservative  55;  Mismatches  90;  Indels   2;  Gaps   1;

Qy    1  MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALELAAAFPGDGWLGSAA   60
         |||||||||| : |||||| :  ||||||||| ||||:||||||||||||||||||||
Db    1  MSGAFIIDPTLKAIEAWHALLGIGVPNDGGVLYSSLSFEKALEHLAAAFPGDGWLGSAA   60

Qy   61  DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT  120
         ||||| ||:| |:|:||| |||:|||||:|||||:||| ||: :||||:||||||:|
Db   61  DKYAGQNRKRVDIFQELAELDKELIELIHNQANSVQTTRGILDGAKKALLFVRPVAIDLN  120

Qy  121  YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD  180
         ||| ||  ||||: |  | |||| |||| ||| : ::|| : :|:|| |||:::|:|
Db  121  YIPLVGSVMSASIQAQACAAAMAAVSGGLAYLLVVLVQTAIHTAKFVALLARLAHLLASAVAD  180

Qy  181  IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT  240
         :|| |:  : :::|::|||: |||:|| :  |: |: |||||||||: :|||| ||||||
Db  181  VVSDGVAIIKGIVDHLWHFIAGALTGLKDIVEKIIHWFFGLFSHWSRLHSFFGGIPGLS  240

Qy  241  GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA  300
         ||||||||||||||  |:|||||| :|:||:: :|||| || |  || |||| ||:|||
Db  241  GATSGLSQVTGLFGVPGLAGSSGLLSGESLLSTENLPSLAGVGAGLGLGSLPQLAQLHAA  300

Qy  301  STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQMGGPVGMGGMHPSSGASK--GTTTK  358
         ||||  | : || | || |||||| : |||| ||||| |||||||| |:: || |:
Db  301  STRQGTRSQAGVSAELSTEQFGGQQEPVSAQSQGMGGSQGMGGMTPASTKSKKDERKKK  360

Qy  359  KYSEGAAAGTEDAERAPVEADAGGGQKVLVRNVV  392
         ||||||||||: :|||||:|| |:::||| ||
Db  361  KYSEGAAAGTDDDAERAPIEVQSSGGGKRALAQHVV  394
```

RESULT 6
US-09-820-843A-18
; Sequence 18, Application US09820843A
; Publication No. US20030039963A1
GENERAL INFORMATION:
; APPLICANT: Council of Scientific and Industrial Research
; TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
; TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 256
; TYPE: PRT
; ORGANISM: M. tuberculosis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: PE_PGRS
; NAME/KEY: misc_feature
; OTHER INFORMATION: gi|3261822

US-09-820-843A-18

Query Match
Best Local Similarity   30.6%;   Score 140.5;   DB 11;   Length 256;
Matches   57;   Conservative   11;   Mismatches   77;   Indels   11;   Gaps   7;

Qy 219 TGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSA-SLP 277
              ::: |||::|:   :   ::| | ::   :
Db   82 TGGFAQGG---RGGFGG-NGNTGASGGMG--GAGGAGGAGGAGGLLIGDGGAGGAGGGIG 134

Qy 278 ALAGIGGGSGFGGLPSLAQVHA----------ASTRQALRPRADGPVGAAA 318
              | | |:|| |||||
Db 135 GAGGVGGGGGGAGGTGGGGVASAFGGGNAFGGRGGRGGDGDGGDGGTGGGAGGAGGAGGAG 194

Qy 319 EQVGGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKYSEGAAAGTEDAERAPVEA 378
              | :| |: :|| |||||
Db 195 GWLSGHSGAHGAMGSGHGEGAGGGGGAGRGEAGEAGAGGGTST--------GTN------PGKA 240

Qy 379 DAGGGQ 384
              |
Db 241 GAPGTQ 246

RESULT 7
US-10-259-678-727
; Sequence 727, Application US/10259678
; Publication No. US20030198974A1
; GENERAL INFORMATION:
; APPLICANT: Cole, Stewart
; APPLICANT: Buchrieser-Brosch, Roland
; APPLICANT: Gordon, Stephen
; APPLICANT: Billault, Alain
; TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
; TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
; TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
; FILE REFERENCE: 3495-0169
; CURRENT APPLICATION NUMBER: US/10/259,678
; CURRENT FILING DATE: 2002-09-30
; PRIOR APPLICATION NUMBER: US/09/060,756
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 727
;   LENGTH: 318
;   TYPE: PRT
;   ORGANISM: Mycobacterium sp.
US-10-259-678-727

Query Match
Best Local Similarity   27.5%;   Score 135;   DB 12;   Length 318;
Matches   58;   Conservative   11;   Mismatches   104;   Indels   38;   Gaps   7;

Qy 191 GTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVT 250
              : :| |:| |:: | ::|| |||||: | |||
Db   14 GTPGAVGSGATGAPG---------------GWLLGDGGAGGSGAAG--SGAPGGAGGAAGLWGTG 61

Qy 251 GLFGAAGLSASSG------------LAHADSLASSASLPALAGIGGGSGFGGLPSLAQVH 298
              || | ::| | :| | |: :| | | | | | | |
Db   62 GAGGSSAGGGGAGGAGGAGGWLLGDGGAGGAGGWLLGDGGASTVLGGTGGGGVGGLWGAGGAG 121

Qy 299 AASTRQALRPRADGPVGAAAEQVGGGQSQLVSAQGSQG-----MGGPVGMGGMHPSSGAS 352
              | | | | | :| ::| | |||||
Db 122 GAGGTGLV--GGDGGAGGAGGTGGLLAGLIGAGGGHGGTGGLSTNGDGGVGG---AGGN 175

Qy 353 KGTTTKKYSEGAAAGTEDAERAPVEADAGGG 383
              || || || | | | | | ||| ||
Db 176 AGMLAGPGGAGGAGG--DGENLDTGGDGGAG 204

RESULT 8
US-09-964-662-1
; Sequence 1, Application US/09964662
; Publication No. US20030166846A1

GENERAL INFORMATION:
; APPLICANT: PROTEIN SPECIALTIES LTD.
; APPLICANT: HSC RESEARCH AND DEVELOPMENT LIMITED PARTNERSHIP
; TITLE OF INVENTION: SELF-ALIGNING PEPTIDES MODELED ON HUMAN ELASTIN AND
; TITLE OF INVENTION: OTHER FIBROUS PROTEINS
; FILE REFERENCE: 041082/0112
; CURRENT APPLICATION NUMBER: US/09/964,662
; CURRENT FILING DATE: 2003-05-08
; PRIOR APPLICATION NUMBER: 09/340,736
; PRIOR FILING DATE: 1999-06-29
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 731
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-964-662-1

Query Match
Best Local Similarity   20.4%;   Score 133.5;   DB 12;   Length 731;
Matches   86;   Conservative   35;   Mismatches 136;   Indels 165;   Gaps   15;

Qy   22 GIGIPN--QGGILYSSLEYFEKALELEAAAFPGDGWLGSAADKYAGKNRNHVNFFQELAD   79
              :|:|: |:|:|:
Db 117 GVGLPGVYPGGVL--------------PGARPPGVGVLPGV----------------- 143

Qy   80 LDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPV----AVDLTYIPVVGHALSAAFQ 134
              |:|:| || ||| |||: |:||| |:| ||||:
Db 144 --------------PTGAGVKPKAPGVGAFAGIPGVGPFGGPQPGVPLGY-PIKAPKLPGGYG 192

Qy 135 APFCAGAM------AVVGGA-LAYLVVKTLINATCLLKLLAKLAELVAAAIADIISDVA 186
              | | | || | |::| | |:||:||:| |: |:| :|
Db 193 LPYTTGKLPYCYGPGGVAGAGKAGYPTGTGVGPQAAAAAKAAKFGAGAAGVLPGVG 252

Qy 187 DIIKGTLGEWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGL 246
              :|||:| :| |:| ||||||||: | |||
Db 253 G--AGVPGVPGA---- 262

Qy 247 SQVTGLFGAAGLSASSGLAHADSLASSALPALAG-IGGGSGFG------GLPSL- 294
              || || : | | | | |: | | | | | : || |||:|
Db 263 --IPGIGGIAGVGTPAAAAAAAAAAKNAKYGAAAGLVPGGPGFGPGVVGVPGAGVPGVGV 320

Qy 295 --------------AQVHAASTRQALRPRADGPVGAAAEQVG---------------- 322
              ||| | |||| |
Db 321 PGAGIPVVPGAGIPGAVPGVVSPEAAAKAAAKAAKYGARPGVGVGGIPTYGVGAGGFPG 380

Qy 323 ---GQSQLVSAQGSQGMGGPVGMG-----GMHPSSGASKGTTTKKYSEG--AAAGTEDAE 372
              | | | | |:: |:|| |:|| | | | | | | |
Db 381 FGVGVGGIPGVAGVPGVGGVPGVGGVPGVGISPEAQAAAAAKAAXYGVGTPAAAAKAAA 440

Qy 373 RA 374
              :|
Db 441 KA 442

RESULT 9
US-10-156-761-13993
; Sequence 13993, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30

```
RESULT 10
US-09-820-843A-20
; Sequence 20, Application US/09820843A
; Publication No. US20030039963A1
; GENERAL INFORMATION:
;  APPLICANT: Council of Scientific and Industrial Research
;  TITLE OF INVENTION: A COMPUTATIONAL METHOD FOR THE IDENTIFICATION OF CANDIDATE PROTEI
;  TITLE OF INVENTION: USEFUL AS ANTI-INFECTIVES
; FILE REFERENCE: Q63915
; CURRENT APPLICATION NUMBER: US/09/820,843A
; CURRENT FILING DATE: 2001-03-30
; NUMBER OF SEQ ID NOS: 118
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 1079;
; TYPE: PRT
; ORGANISM: M. tuberculosis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: PE_PGRS
; NAME/KEY: misc_feature
; OTHER INFORMATION: gi|2924449
US-09-820-843A-20

Query Match             6.6%;   Score 130;  DB 11;   Length 1079;
Best Local Similarity  26.4%;   Pred. No. 0.016;
Matches   46;  Conservative  12;  Mismatches  72;  Indels  44;  Gaps  4;

Qy    234 AGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPA------------ 278
              ||:||: ||:||:|||||||||||:|||||: ||: :||:
Db    313 AGAGGAAGSNGGTVGANGTGGDGNGGAAGAATAGSNGGAGTGSAGGNGGTGGRGGSGGA 372

Qy    279 ----LAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGG 334
              :||||| ||:: :||||:||||
Db    373 GDGIGGVGGGKKGGNG------------ADGEVGGAG-GAGGSGPNTSPGGNG 412

Qy    335 GMGGPVGMGGMHPSSGA----SKGTTTKKYSEGAAAGTEDAERAPVEADAGGG 383
              || |||||||||:|    :: : :||| :::||||
Db    413 GQGQGGGSGSGAGAAGAGAGGANGTAGNGGQGGAGGGTGGAGAASSATNGGSG 466
```

```
RESULT 11
US-09-961-403-8
; Sequence 8, Application US/09961403
; Publication No. US2003007589A1
; GENERAL INFORMATION:
;  APPLICANT: HE-STUMPP, HOLGER
;  APPLICANT: HAENDLER, BERNARD
;  APPLICANT: KRAETZSCHMAR, JOERN
;  APPLICANT: KREFT, BERTHOLT
;  APPLICANT: WINTERHAGER, ELKE
;  APPLICANT: REGIDOR, PEDRO
;  APPLICANT: SCOTTI, SIMONE
;  TITLE OF INVENTION: METHOD FOR IN VITRO DIAGNOSIS OF ENDOMETRIOSIS
; FILE REFERENCE: SCH-1789
; CURRENT APPLICATION NUMBER: US/09/961,403
; CURRENT FILING DATE: 2001-09-25
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 730;
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-961-403-8

Query Match             6.6%;   Score 129.5;  DB 11;   Length 730;
Best Local Similarity  20.3%;   Pred. No. 0.01;
Matches   86;  Conservative  36;  Mismatches 134;  Indels 167;  Gaps  15;

Qy     22 GIGIPN--QGGILYSSLEVFEKALEELAAAFPGDGWLGSAADKYAGKNRNHVNFFQELAD 79
              |::|: :|::|
Db    143 GVGLPGVYPGGVL-----------------PGARFPGVGVLPGV-------------- 169

Qy     80 LDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPV----AVDLTYIPVVGHALSAAFQ 134
              |: :::
Db    170 --------------PTGAGVKPKAPGVGGAFAGIPGVGPFGGPQPGVPLGY-PIKAPKLPGGYG 218

Qy    135 APFCAGAM------------AVVGGA-LAYLVVKTLINATQLLKLLAKLAELVAAAIADIISDVA 186
              |:|| :::
Db    219 LPYTTGKLPYGYGPGGVAGAGKAGYPTGTGVGPQAAAAAKAAAKFPGAGAGVLPGVG 278

Qy    187 DIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGL 246
              |||
Db    279 G--------------------------------------AGVPGVPGA---- 288

Qy    247 SQVTGLFGAAGLSASSGLAHADSLASSASLPALAG-IGGGSGFG------GLPSL- 294
              ||:|: |||| ||: :: :: ||||:|
Db    289 --IPGIGGIAGVGTPAAAAAAAAAAAAKYGAAAGLVPGGPGFGPGPGVGVGVPGAGVPGVGV 346

Qy    295 --------AQVHAASTRQALRPRADGPVGAAAEQVG--------------------- 322
                           :|:
Db    347 PGAGIPVVPGAGIPGAAVPGVVSPEAAAKAAAKAAKYGARPGVGVGGIPTYGVGAGGPG 406

Qy    323 -----------GQQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKYSEG--AAAGTEDA 371
                         ||:||: |||
Db    407 FGVGVGGIPGVGAGVPSVGGVGVGGVPGV-GISPEAQAAAAKAAKYGVGTPAAAAAKAA 465

Qy    372 ERA 374
              :|
Db    466 AKA 468
```

```
RESULT 12
US-10-097-111-283
; Sequence 283, Application US/10097111
; Publication No. US20030138771A1
; GENERAL INFORMATION:
;  APPLICANT: PELLETIER, JERRY
;  APPLICANT: GROS, PHILLIPPE
;  APPLICANT: DUBOW, MICHAEL
;  TITLE OF INVENTION: DNA SEQUENCES FROM S. PNEUMONIAE BACTERIOPHAGE EP1 THAT
;  TITLE OF INVENTION: ENCODE ANTI-MICROBIAL POLYPEPTIDES
```

```
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 13993
; LENGTH: 446
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-13993

Query Match             6.7%;   Score 131.5;  DB 15;   Length 446;
Best Local Similarity  25.7%;   Pred. No. 0.0034;
Matches   74;  Conservative  33;  Mismatches 112;  Indels  69;  Gaps  11;

Qy     94 AVQTTRDILEGAKKGLEFVRPVAVDLTYIPVVGHALSAAFQAPFCA---GAMAVVGGAL- 149
              | :: :|| :|: |: :  | :|||| |
Db     51 AITSLGALLEVRDGLGMSGSVAGLLTSVPPLCFAVFGV-MAPRLARRFGAAAVVGAGMV 109

Qy    150 ---AYILVVKTLINATQLLKLLAKLAELVAAAIADIISDVADIIKGTLGEVWEFITNALNG 206
              :|||:
Db    110 AITAGLVIRPYAGGTAGF-LAASALALMGIAVSNVLMPVI---------------- 148

Qy    207 LKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVT-GLFGAAGLSASSGLA 265
              :: :||||                              ||||
Db    149 VKRWFPDRVGSMTGLYSM---------ALALGTSAAAVTVPLTDALGGSWQSGLA 195

Qy    266 HADSLASSASLP------ALAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAE 319
              ||:|: ||                            |: :|
Db    196 VWAGLAAAAVLPWIPLVRARGGAGSHGFGSSEAAGGSHGLGSSEA-------AAAAAP 247

Qy    320 QVGGGSQLVSAQ--------GSQGMGGPVGMGGM--HPSSGASKGT 355
              :|:: : :   |     :: | |:|
Db    248 QEADGLRITRSRTAWALAVFGLQATAAYITMGWMAQIFRDAGVAAGT 295
```

```
;       FILE REFERENCE: 073406-0603
;       CURRENT APPLICATION NUMBER: US/10/097,111
;       CURRENT FILING DATE: 2002-07-24
;       PRIOR APPLICATION NUMBER: 09/676,412
;       PRIOR FILING DATE: 2000-09-29
;       PRIOR APPLICATION NUMBER: 60/157,218
;       PRIOR FILING DATE: 1999-09-30
;       NUMBER OF SEQ ID NOS: 552
;       SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 283
;       LENGTH: 1149
;       TYPE: PRT
;       ORGANISM: Streptococcus pneumoniae
US-10-097-111-283

Query Match        6.6%;  Score 129; DB 12; Length 1149;
Best Local Similarity   19.8%;  Pred. No. 0.022; Indels 138; Gaps   16;
Matches   87;  Conservative   64; Mismatches 150; Indels 138; Gaps   16;

Qy       6 IIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAADKYAG  65
              | | ::::   :  :  ||:::::::::::::::::::::::::::::::::::::::::
Db     686 IIQAAVQIITALFNGLVQALPT---LIQAGLQILSALINGLVQALPA-----------  729

Qy      66 KNRRNHVNFFQELADLDRQLI-SLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLTYIPV 124
              ::   :    :: ::::::|   | :::  ::      ::: ::  :
Db     730 ----IIQAAVQIIMSLVQALIENLPMIIBAAMQIIMGLVNAL---------------IEN  770

Qy     125 VGHALSAAFQAPFCAGAMAVVGGALAYL----------VVKTLI-----NATQLLK----  165
              | :   ::  |||   |          | | : |       :: | |   |:|
Db     771 IGPTIEAGIQI-----LMALIEGLIQVLPELITAAIQIITSLLEAILSNLPQLLEAGVKL  825

Qy     166 -------LLAKLAELVAAAI---ADIISDVAD------------IIKGTLGEVWEFI  200
                       ||||::|  |:|    :::                  |||::||:||:
Db     826 LLSLLQGLLNMLPQLIAGALQIQMMALLKAVIDPVPKLLQAGVQLLKALIQGIASLLGSLL  885

Qy     201 TNALNGLKELWLDKLTGWVTGLFSRGWSNLESFFAGVPGLTGA-----------TSGLSQVTG  251
              ::::: ::::| ::::  : :|    | : ||    :|   |             ::::::||
Db     886 STAGNMLSSLVSKIASFVGQMVSGGANLIRNFISGISMIGSAVSKIIGSMGTSIVSKVTG  945

Qy     252 LFG--AAGLSASSGLAHADSLASSASLPALAGIGGGS-----GFGGLPSLAQVHAASTR  303
              | |  :::::::::||:|:|:  ::::    |            |:|::::::   ::
Db     946 FAGQMVSAGVNLVRGFINGISSMVSSAVSAAANMASSALNAVKGFLGIHSPSRV-------  999

Qy     304 QALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKYSEG 363
              :: :| : | |:::::   ||::::||:::::  ::: |::  |:::::::   ::
Db    1000 ------------MEQMGIYT-----------GGQFVNGIGNM-------IRTTRDKAKEM 1029

Qy     364 AAAGTEDAERAPVEADAGG 382
              ::  :|:::::: : ::
Db    1030 AETVTEALSDVKMDIQENG 1048

RESULT 13
US-09-738-626-6032
;  Sequence 6032, Application US/09738626
;  Publication No. US20020197605A1
;  GENERAL INFORMATION:
;  APPLICANT: NAKAGAWA, SATOSHI
;  APPLICANT: MIZOGUCHI, HIROSHI
;  APPLICANT: ANDO, SEIKO
;  APPLICANT: HAYASHI, MIKIRO
;  APPLICANT: OCHIAI, KEIKO
;  APPLICANT: YOKOI, HARUHIKO
;  APPLICANT: TATEISHI, NAOKO
;  APPLICANT: SENOH, AKIHIRO
;  APPLICANT: IKEDA, MASATO
;  APPLICANT: OZAKI, AKIO
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-125
;  CURRENT APPLICATION NUMBER: US/09/738,626
;  CURRENT FILING DATE: 2000-12-18
;  PRIOR APPLICATION NUMBER: JP 99/377484
;  PRIOR FILING DATE: 1999-12-16
```

```
;       PRIOR APPLICATION NUMBER: JP 00/159162
;       PRIOR FILING DATE: 2000-04-07
;       PRIOR APPLICATION NUMBER: JP 00/280988
;       PRIOR FILING DATE: 2000-08-03
;       NUMBER OF SEQ ID NOS: 7059
;       SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 6032
;       LENGTH: 585
;       TYPE: PRT
;       ORGANISM: Corynebacterium glutamicum
US-09-738-626-6032

Query Match        6.4%;  Score 127; DB 10; Length 585;
Best Local Similarity   21.0%;  Pred. No. 0.013;
Matches  115;  Conservative   54; Mismatches 186; Indels 192; Gaps   24;

Qy      12 SAIDGLYDLLGIG----IPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAADKYAGKN  67
              : : :|  : |    ::::::::::::::::::::::::::::::::::::::::::
Db      41 SPVSGL-DQLGSGHGNVINGGAGAANSVLNSYAEQIEWLSAALQASGAALTGQDELFARG  99

Qy      68 RNHVNFFQELADLDRQLIS---------------LIHDQANAVQTTRDIL----EGAKK 107
              :   :                           ::
Db     100 ------MDVADTGGRVVEESVMPPARPAPRFESFVFNPPAVSPSLSDLCSQPSGTNS 152

Qy     108 G----------LEFVRPVAVDLTYIPVVGHALS---------------AAFQAPFC 138
              |                ::::
Db     153 GAVLEAQGSWGSMASAISNVSASLS--STAGEILAENSGETFEQAAARINEVAAAGATFA 210

Qy     139 AGAM---AVVGGA------------LAYLVVKTLINATQLLKL-LAKLAELVAAAIADI 182
                    :::|                ::   :   :  |  ::
Db     211 ANAKMMGASVGTLNRIYMGHRMQVFMAATSIKAILDPVQRLAABRAFLASFQATFQADVL 270

Qy     183 SDVADI--------IKGTLGEVWEFI-TNALNGLKElWDKLTGWVTGLFSRGWSNLESFF-- 233
              ::                 ::   |  :: ::::::::: :: : ||
Db     271 TGMPPVSNLMQMKGANGSAGEIALGMDEIAGSGQAWSAAGLTPSGAAQGGVANAGSIAPD 330

Qy     234 AGVPGLTGATSGLSQVTGLFG-----------AAGLSASSGLAHADSLASSA 274
              ::|        : : |             :::::::::::::|:|:|:
Db     331 AAVQGAGQSG------VGSFGTVTDQLDGINIGDMLTSAASAGQSLANGLAMPTSTPNSA 385

Qy     275 S--LPA--------LAGIGGGSGFG----------------------------GLP 292
              :  ||:        ::
Db     386 SGAIPSSMSAASPLGAFGSGAGLGAQGGSIGSSAPGAISSRAAGSAGGSVPGMTGGPGAP 445

Qy     293 SL----AQVHAASTRQALRP-----------RADGPVGA----------AAEQVGGQSQ 326
              :      ::
Db     446 GITSDSLMGARTHGASSAGAVAPMMGGAGGMSGGVVGAGGTGSQSKYARQTGSSVGSSSQ 505

Qy     327 LVSAQGSQGMG-------------GPVGMGGMIPSSGASKG-TTTKKYSEGAAAGTEDAERA 374
              ::                      :::   :::: :   ::
Db     506 SGSGLGMVGSGSGSGKPSISNFGRGMMPMMPMMPMPMGGAGGGGQXNTGKVKTVTSAVEEDRNLA 565

Qy     375 PVEADAG 381
              ::::|
Db     566 ALLGDRG 572

RESULT 14
US-10-259-678-728
;  Sequence 728, Application US/1025/9678
;  Publication No. US20030198974A1
;  GENERAL INFORMATION:
;  APPLICANT: Cole, Stewart
;  APPLICANT: Buchrieser-Brosch, Roland
;  APPLICANT: Gordon, Stephen
;  APPLICANT: Billault, Alain
;  TITLE OF INVENTION: METHOD FOR ISOLATING A POLYNUCLEOTIDE OF INTEREST FROM
;  TITLE OF INVENTION: THE GENOME OF A MYCOBACTERIUM USING A BAC-BASED DNA
;  TITLE OF INVENTION: LIBRARY APPLICATION TO THE DETECTION OF MYCOBACTERIA
;  FILE REFERENCE: 3495-0169
;  CURRENT APPLICATION NUMBER: US/10/259,678
;  CURRENT FILING DATE: 2002-09-30
;  PRIOR APPLICATION NUMBER: US/09/060,756
```

```
; PRIOR FILING DATE: 1998-04-16
; NUMBER OF SEQ ID NOS: 743
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 728
;   LENGTH: 334
;   TYPE: PRT
;   ORGANISM: Mycobacterium sp.
US-10-259-678-728

Query Match         6.4%;  Score 126.5; DB 12;  Length 334;
Best Local Similarity   27.6%;  Pred. No. 0.0063;
Matches    55;  Conservative   11;  Mismatches   90;  Indels   43;  Gaps    8;

Qy    191  GTLGEVWEFITNALNGLKELMDKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVT  250
               ||||||||                 |:|  |||         :|::|
Db     14  GTPGAVGSGATGAPG--------GWLLGDGGAGGSGAAG--SGAPGGAGGAAGL--W      58

Qy    251  GLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAASTRQALRPRA  310
               |:|    ||         |:|                 ||||
Db     59  GTGGAGGIGGAS---------------------TVLGTGGDGVGGLWGAGGAGGAGGTGLV--GG  102

Qy    311  DGPVGAAAEQVGGGQSQLVSAQGSQG------MGGPVGMGGMHPSSGASKGTTTKKYSEGA  364
               ||                |||          :|       ||   ||| :
Db    103  DGGAGGAGGTGGLLAGLIGAGGGHGGTGGLSTNGDGGVGG----AGGNAGMLAGPGGGAGG  158

Qy    365  AAGTEDAERAPVEADAGGG  383
               |||
Db    159  AGG--DGENLDTGGDGGAG  175


RESULT 15
US-10-156-761-11729
; Sequence 11729, Application US/10156761
; Publication No. US2003011901A1
; GENERAL INFORMATION:
;   APPLICANT: OMURA, SATOSHI
;   APPLICANT: IKEDA, HARUO
;   APPLICANT: ISHIKAWA, JUN
;   APPLICANT: HORIKAWA, HIROSHI
;   APPLICANT: SHIBA, TADAYOSHI
;   APPLICANT: SAKAKI, YOSHIYUKI
;   APPLICANT: HATTORI, MASAHIRA
;   TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;   FILE REFERENCE: 249-262
;   CURRENT APPLICATION NUMBER: US/10/156,761
;   CURRENT FILING DATE:  2002-05-29
;   PRIOR APPLICATION NUMBER: JP 2001-204089
;   PRIOR FILING DATE: 2001-05-30
;   PRIOR APPLICATION NUMBER: JP 2001-272697
;   PRIOR FILING DATE: 2001-08-02
;   NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11729
;   LENGTH: 749
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-11729

Query Match         6.4%;  Score 126; DB 15;  Length 749;
Best Local Similarity   23.6%;  Pred. No. 0.022;
Matches   109;  Conservative   58;  Mismatches  167;  Indels  128;  Gaps   25;

Qy      3  RAFIDPTISAIDGLYDLLG-IGIPNQGGILYSSLEYFEKALELAAAFPGDGWLG----  57
               |::|||.|| : |   :||:                :||
Db    235  QAFLILPALAA---LYAVCAPVKLKKRLGQL-------ALAGVAMVVAGGWWVAIVEL  282

Qy     58  --SAADKYAGKNRNHVNFFQELADLDRQLISLIHDQANAV------------QTRD  100
             :::     |  |:|            |    |: :|
Db    283  WPASSRPYIGGSQN--NSPLBLTFGYNGLGRINGEETGSVGGGGGGGGGTQGWGETGWN  340

Qy    101  ILBGAKKG--LEFVRPVAVDLTYIPVGHALS-AAFQAPFCAGAMAVVGGALAY--LVVK  155
               |:    ::| :||| | |  |||:         |||        :|
Db    341  RMFNSEIGGQISWLLPAAILF--VAGLWLTRKAKRTDLVRGSFLAWGGALLMTGLIFS  397
```

Now continuing (upper-right block, SEQ ID NO 728 alignment continued):

```
Qy    156  TLINATQLLKLLAKLAELVAAAIADIISDVADIIKGTLGEVWEFITNA--------  203
               |:|     :|   :|::|  :|||
Db    398  FMAGIFHQYYTVA-----LAPYIAALIGMGATVLWEERTKVWASLTLAGAVVASAVWGYV  452

Qy    204  -LNGLKEL-----WDKLTGWVTGLFSRGWSNLESFFAGVPG--LTGATSGLSQVTGLFGA  255
               ||  |              :|            :|||||||
Db    453  LLNRTSDYLPWLKWFVLVGGLAGA-------LGLVFAAKLGRRLALAAVGLSFVASVAGP  505

Qy    256  AGLSASS-GLAHADSLASSASLPALAGI--------GGGSGF-GGLPSLAQVH---AAST  302
               :|:||  :|  :|:                 |||||  ||||||||        :|
Db    506  TAYILSTVNTGHTGSIVTAG--PAGAGMGRGGPGGGGGMRGGFPPGAGGNQQGGTGNT  563

Qy    303  RQALRPRADGPVGAAAEQ----VGGQSLVSAQGSQ--------GMGG  338
               |||:                  ||:            ||||
Db    564  QQGGGPQGGGAGGGMGQPPTGGFGNQQNGNAQGQNQQGMPGGGQTGERGGMGG  623

Qy    339  PVGMGGMHPSSGASKGTTTKKYSE-------GAAAGTEDA  371
               |||   |||||||||||   |:         :|:
Db    624  --GMGGL--LSGASVSSAAKKLLETDADSYTWAAAAIGAQNA  661


Search completed: November  5, 2003, 17:07:22
Job time : 47 secs
```

This Page Blank (uspto)

```
                 GenCore version 5.1.6
       Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on:  November  5, 2003, 16:59:57 ; Search time 29 Seconds
                                       (without alignments)
                              1299.935 Million cell updates/sec

Title:         US-09-688-672A-14
Perfect score: 1969
Sequence:      1 MSRAFIIDPTISAIDGLYDL........RAPVEADAGGGQKVLVRNVV 392

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:      283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters:      283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     PIR 76.*
               1: pir1:*
               2: pir2:*
               3: pir3:*
               4: pir4:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Match Length	DB	ID	Description
1	1969	100.0	392	2	A70957	hypothetical prote
2	1258	63.9	394	2	S72814	hypothetical prote
3	380.5	19.3	402	2	E70656	hypothetical prote
4	168	8.5	496	2	H70839	hypothetical glyci
5	155	7.9	1329	2	E70917	hypothetical glyci
6	153	7.8	837	2	S30971	gene 26 protein -
7	152	7.7	801	2	F70824	hypothetical glyci
8	151	7.7	741	2	G70917	hypothetical glyci
9	150	7.6	1306	2	A70934	hypothetical glyci
10	149.5	7.6	916	2	T03323	gene 116 protein -
11	148.5	7.5	515	2	H70663	hypothetical glyci
12	148.5	7.5	749	2	A70812	hypothetical glyci
13	148	7.5	409	2	A70647	probable PPE prote
14	148	7.5	588	2	F70971	hypothetical glyci
15	148	7.5	957	2	D70835	hypothetical glyci
16	148	7.5	1660	2	A70869	hypothetical glyci
17	147	7.5	439	2	D70954	hypothetical glyci
18	146.5	7.4	837	2	E70835	hypothetical glyci
19	145.5	7.4	1538	2	H70846	hypothetical glyci
20	145	7.4	914	2	H70987	hypothetical glyci
21	144.5	7.3	539	2	A70899	probable PPE prote
22	144	7.3	1381	2	E70806	hypothetical glyci
23	142.5	7.2	525	2	D70878	hypothetical glyci
24	142.5	7.2	667	2	A70893	hypothetical glyci
25	142.5	7.2	882	2	B70812	hypothetical glyci
26	142	7.2	461	2	F70571	hypothetical glyci
27	142	7.2	778	2	F70963	hypothetical glyci
28	141.5	7.2	1489	2	D70807	hypothetical glyci
29	141	7.2	584	2	G70804	hypothetical glyci
30	141	7.2	837	2	H72802	minor tail subunit
31	140.5	7.1	256	2	A70514	hypothetical glyci
32	140.5	7.1	853	2	A70896	hypothetical glyci
33	139.5	7.1	543	2	F70726	hypothetical glyci
34	137.5	7.0	783	2	E70824	hypothetical glyci
35	137.5	7.0	1901	2	F70806	hypothetical glyci
36	136	6.9	358	2	H70857	probable PPE prote
37	136	6.9	618	2	A70989	hypothetical glyci
38	136	6.9	731	2	C70974	hypothetical glyci
39	135.5	6.9	457	2	H70820	hypothetical glyci
40	135	6.9	923	2	E70820	hypothetical glyci
41	135	6.9	1517	2	T13329	hypothetical prote
42	134.5	6.8	361	2	G70682	hypothetical glyci
43	134	6.8	513	2	D70836	probable PPE prote
44	134	6.8	606	2	H70816	hypothetical glyci
45	133.5	6.8	940	2	F84089	phage-related prot

ALIGNMENTS

```
RESULT 1
A70957
hypothetical protein Rv3616c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: A70957
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
;Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70957
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-392 <COL>
A;Cross-references: GB:Z95557; GB:AL123456; NID:g3242276; PIDN:CAB08950.1; PID:e316833;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3616c

Query Match    100.0%;  Score 1969;  DB 2;  Length 392;
Best Local Similarity  100.0%;  Pred. No. 8.9e-119;
Matches  392;  Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPDGWLGSAA 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPDGWLGSAA 60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLLHDQANAVQTRDILEGAKKGLEFVRPVAVDLT 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DKYAGKNRNHVNFFQELADLDRQLISLLHDQANAVQTRDILEGAKKGLEFVRPVAVDLT 120

Qy  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLLAELVAAAIAD 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLLAELVAAAIAD 180

Qy  181 IISDVADIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  181 IISDVADIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240

Qy  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSPFGLPSLAQVHAA 300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSPFGLPSLAQVHAA 300

Qy  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGPVGMGMHPSSGASKGTTTKKY 360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301 STRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGPVGMGMHPSSGASKGTTTKKY 360

Qy  361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
          |||||||||||
```

Db   361 SEGAAAAGTEDAERAPVEADAGGGQKVLVRNVV 392

RESULT 2
S72814
hypothetical protein B1620_C2_213 - Mycobacterium leprae
C;Species: Mycobacterium leprae
C;Date: 19-Mar-1997 #sequence_revision 25-Apr-1997 #text_change 23-Mar-2001
C;Accession: S72814
R;Smith, D.R.; Robison, K.
submitted to the EMBL Data Library, November 1993
A;Description: Mycobacterium leprae cosmid B1620.
A;Reference number: S72584
A;Accession: S72814
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-394 <SMI>
A;Cross-references: EMBL:U00015; NID:g466931; PIDN:AAC43223.1; PID:g466935

Query Match   63.9%;   Score 1258; DB 2; Length 394;
Best Local Similarity 62.7%; Pred. No. 2.7e-73;
Matches 247; Conservative 55; Mismatches 90; Indels 2; Gaps 1;

Qy   1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA 60
          ||| ||||||||:||||||||||||||:|||||||||||||| ||||||||||||||||||
Db   1 MSGAFIIDPTIKAIEAWHALLGIGVPNDGGVLYSSLSFFEKALEHLAAAFPGDGWLGSAA 60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTRDILEGAKKGLEFVRPVAVDLT 120
          ||||||||| |||||||||||||| :||| :|||||| ||||| ||||||||||||||
Db   61 DKIAGQNRKRVDIFQELAELDKEIIELIHNQANSVQTRGILDGAKKALLFVRPVAIDLN 120

Qy   121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLABLVAAAIAD 180
          |||| | |||:|||||||||||||||:|||||| :||||||||||:| :|||||||||||
Db   121 YIPLVGSVMSASIQAQACAAAMAAVSGGLAYLLVQTAIHTAKPVALIARLAHLLASAVAD 180

Qy   181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLT 240
          :||||||||||:||||||| ||| |||| ||||||||||| |||||||||||||||||||
Db   181 VVSDGVAIIKGIVDHLWHFIAGALTGLKDIVEKIIHWPFGLFSHWWSRLHSFFGGIPGLS 240

Qy   241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
          |||||||||||||||||| :||:| ||||:|||| |::|| | ||||:||| :||||||||
Db   241 GATSGLSQVTGLFGVPGLAGSSGLLSGESLLSTENLPSLAGVGAGLGLGSLPQLAQLHAA 300

Qy   301 STRQALRPRADGPVGAAEBQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASK--GTTTK 358
          ||| || |:||||||||: |||||||||||||||||||||||||:|||||| ::::| |
Db   301 STRQGTRSQAGVSAELSTEQFGGGQQEPVSAQGSQGMGGSQGMGGMTPASTKSKKDERKKK 360

Qy   359 KYSEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
          ||||||||||||||||||||||||||| :| :||
Db   361 KYSEGAAAGTDDAERRAPIEVQSGGGKRALAQHVV 394

RESULT 3
E70656
hypothetical protein Rv3864 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: E70656
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: E70656
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-402 <COL>
A;Cross-references: GB:Z83864; GB:AL123456; NID:g3261687; PIDN:CAB06237.1; PID:e301440;
C;Genetics:
A;Gene: Rv3864

Db   361 SEGAAAAGTEDAERAPVEADAGGGQKVLVRNVV 392

Query Match   19.3%;   Score 380.5; DB 2; Length 402;
Best Local Similarity 30.1%; Pred. No. 3.6e-17;
Matches 129; Conservative 47; Mismatches 157; Indels 95; Gaps 13;

Qy   10 TISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAADKYAGKNRN 69
          |:| | | : | | | || | |:| | | |: |: | | :|
Db   11 TSNFIWQLLLLGEGIPDPGDIFNTGSSLFKQISDRMGLAIPGTNWIGQAAEAYLNQNIA 70

Qy   70 HVNFFQELADLDRQLISLIHDQANAVQTRDILEGAKKGLEFVRPVAVDLTYIPVVGHAL 129
          |:|| || || | :||| | | :|: | |:| || | :| | ||| |:|| ||| |
Db   71 QQLRAQVMGDiDKLTGNMISNQAKVVSDTRDVLRAMKKMIDGVYKVCKGLEKIPLLGHLW 130

Qy   130 SAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIADIISDVADII 189
          | |: |:| ||||||| ||| | || ||||||| |: ||
Db   131 SWELAIPMSGIAMAVVGGALLYLTIMTLMNATNL----------------------- 164

Qy   190 KGTLGEVWEFITN-----ALNGLKELWDKLTG-W------------------VTGLFSRGW- 226
          | :   :| || :||| :   : :
Db   165 RGILGRLIEMLITTLPKFPGLPGLSLPDIIDGLWPPKLPDIPIPGLPDIPGLPDFKWPPT 224

Qy   227 -------SNLESP----------FAGVPGLTGATSGLSQVTGLF-GAAGL----------- 258
                    :| || | | | : : | ||| || |
Db   225 PGSPLFPDLPSPSPGPGFPGFPEFPAIPGFP-ALPGLESIPNLFPGLPGLGDLLPGVGDLGKL 283

Qy   259 -SASSGLAHADSLASSASLPALAGIGGGSFGGLPSLAQVHAASTRQALRPRADGPVGAA 317
                :| | : | | | : ||: | | :   :| || || | |: | | |
Db   284 PTWTELAALPDFLGGFAGLPSL-GFGNLLSPASLPTVGQVTATMGQLQQLVAAGGGPSQL 342

Qy   318 AEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKYSEGAAAGTEDAERAPVE 377
          | : || | :| : | | |: | : |
Db   343 ASMGSQQAQLISSQAQGG------------GQQHATLVSDK----KEDEGVA-------- 387

Qy   378 ADAGGGQK 385
          | | :|
Db   388 AGTAASQR 395

RESULT 4
H70839
hypothetical glycine-rich protein Rv0109 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: H70839
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome s
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70839
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-496 <COL>
A;Cross-references: GB:AL021926; GB:AL123456; NID:g3261520; PIDN:CAA17303.1; PID:g2909574
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0109
C;Superfamily: Phaseolus glycine-rich cell wall protein 1.8

Query Match   8.5%;   Score 168; DB 2; Length 496;
Best Local Similarity 25.5%; Pred. No. 0.0018;
Matches 82; Conservative 25; Mismatches 113; Indels 102; Gaps 13;

Qy   115 VAVDLTYIPVVGHALS---AAFQAPFCAGAMAVVG---------------- 146
               | | | |: | ||| | |||| |
Db   11 VAAAATHLAGIGSALSTANAAAAAP--TTALSVAGADEVSVLIAALFEAYAQEYQALSAQ 68

Qy   147 ------------GALAYLVVKTLINATQLLKLLAKLAELVAAAIADIISDVADIIKGT 192
                       || |||:| ||| |||||| ||||| || |||||| :||||||: |
Db   69 ALAFHDQFVQALNMGAVCYAAAET-ANATPIQAL-----QTVQQNVLTVNAPTQALLGR 122

A;Title: Superinfection immunity of mycobacteriophage L5: applications for genetic trans
A;Reference number: S30949; MUID:9321283; PMID:8459767
A;Accession: S30971
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-837 <DON>
A;Cross-references: EMBL:Z18946; NID:g15859; PIDN:CAA79402.1; PID:g579124
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1992
C;Genetics:
A;Gene: 26
A;Start codon: GTG

Query Match   7.8%; Score 153; DB 2; Length 837;
Best Local Similarity   22.3%; Pred No. 0.03;
Matches 101; Conservative 59; Mismatches 157; Indels 136; Gaps   18;

Qy   13 AIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELEAAAFPGDWGLGSAADKYAGKNRNHVN 72
            :| ::| : |:
Db   310 AMQGLSQTLG-------------SVLNLFNRLMESGLQAM---GQLG-------GPLSTFIN 348

Qy   73 FFQE------------------LADLDRQLISLIHDQANAVQTRDILEGAKKG-LEFV 112
            ||| : | :|
Db   349 GFGDLFVSLMPALTSVSGLIGNVLGTLGTQLAPIVTALTPAFQTLASTLGTMLTGALQAL 408

Qy   113 RPVAVDLTVI-PVVGHALSAAFQA------------------------PFCAGAMAV 144
           | || | ||
Db   409 GPI---LTQVATLIGTTLNTALQLQPMLPSLMQSFQQISDVLVTSLAPHIPALATALGQ 465

Qy   145 VGGA---LAYLVVKTLINATQLLKLLAKLAEL------------------V 174
           ||| :|:|: :|||
Db   466 VAGAVLQLAPTIISTLVPA--FVQLVPKVAELVPTIVNLVQSPANLMPVVLPLAQALVSV 523

Qy   175 AAAIADIISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLSRGWSNLESFFA 234
            ::   ::|: |:|
Db   524 AGAVIQVGVSIGGALIGALANLTELISNVIK------KVSEWVSS-FSSGAQQIAAKAA 575

Qy   235 GVPGL------------TGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGG 284
           |||| :: |:   ::|| :
Db   576 ELPGMIQSALANLMAIG|QAGKDLVQGLINGIGMVSAAVNKAKELASSVA-------GA 628

Qy   285 GSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGGQSQLVSAQGSQGMGGPVGMGG 344
           |:   : ::   :: |::| |: || ::   ::|:|
Db   629 VKGPFLGIESPSKLFTEYGQPTAEGFGNG-MEAGFKPVIERAKDLAAELSRAM-----ESG 682

Qy   345 MHPSSGASKGTTTKKYSEGAAAGTEDAERAPVE 377
           |: :   :   ::| : :
Db   683 TDP-SGILAGLDQNELKQMLAALEEERKRLKVE 714

RESULT 7
F70824
hypothetical glycine-rich protein Rv0747 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: F70824
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
; Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70824
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-801 <COL>
A;Cross-references: GB:AL021958; GB:AL123456; NID:g3261536; PIDN:CAA17514.1; PID:g291102
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0747
C;Superfamily: elastin

'Query Match   7.7%; Score 152; DB 2; Length 801;

Qy   193 LGEVWEFITNALNGLKELWLDKLTGWTG----LFSRGWSNLESFFAGVPGLTGATSGLSQ 248
            :| | :|:|
Db   123 -----PIIGNGANGLPN------TGQDGGPGGLLFGNGGN-------------GGSGGVDQ 159

Qy   249 VTGLFGAAGL--SASSGLAHADSLASSASLPA------LAGIGGGSGFGGLPSLAQVHAAS 301
            |:||| ||| : |
Db   160 AGGNGAAGLLIGNGGSGSGVGGPGIAGSAGGAGGLLFGNGGPGGAGGIGTTGD----- 214

Qy   302 TRQALRPRADGPVGAAAEQVGGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKYS 361
            |||| :| | | | |:
Db   215 --------GGPGGAGGNAIG----LFGSGGTGGMCGVGGMGGVGNGGVGNGGTAGLFG 261

Qy   362 EGAAAGTEDAERAPVEADAGGG 383
            | ::
Db   262 HGGAGGAGGIGSADGLGGGGGG 283

RESULT 5
E70917
hypothetical glycine-rich protein Rv1450c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: E70917
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: E70917
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1329 <COL>
A;Cross-references: GB:Z95844; GB:AL123456; NID:g3250713; PIDN:CAB09271.1; PID:g2131046
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1450c
C;Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology

Query Match   7.9%; Score 155; DB 2; Length 1329;
Best Local Similarity   26.3%; Pred. No. 0.039;
Matches 79; Conservative 27; Mismatches 132; Indels 62; Gaps 13;

Qy   110 EFVRPVAVDLTYIPVVGHALSAAFQAPFCAGAMAVV--GGALALVLVVKTLIN------- 159
            || |:   :|||
Db   9 ETVAAAALDVARI---GSSIGAANAA--AAGSTTSVLAAGADEVSAAIATLFGSHAREYQ 63

Qy   160 --ATQLLKLILAKLAELVAAAIADIISDVADIIKGTLGEVWEFITNALN-------GLKEL 210
            :||| |:|::|||
Db   64 AISTQVAAFHDRFAQTLSAAVGSYVSAEA-TNAAPLATLEHNVLNALNAPTQALLGRPLI 122

Qy   211 WDKL-----TGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLA 265
            |||| :|:| |||
Db   123 GDGAAGAPGTCQAGGAGGILWGNGGAGGSGAPGQVGGAGG--AAGLFGTGGAGGAGGAG 179

Qy   266 HADSLASSASLPALAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQS 325
            || |:
Db   180 AAGGAGGSGGGW--LLGNGGYGGAGG-----------QSL------LGGATGGAGGNA 217

Qy   326 QLVSAQGSQGMGGPVGMGGMHPSSGAS--KGTTTKKYSEGAAAGTEDAERAPVEADAGGG 383
            :| ::| | ::
Db   218 GLFGVGTGGPGGPGGPGVGTGGAGGLGGTLYGAGGHGGAGGP------GPIGGVGGHG 272

RESULT 6
S30971
gene 26 protein - Mycobacterium phage L5
C;Species: Mycobacterium phage L5
C;Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 20-Jun-2000
C;Genetics:
R;Donnelly-Wu, M.K.; Jacobs Jr., W.R.; Hatfull, G.F.
Mol. Microbiol. 7, 407-417, 1993

Best Local Similarity 28.1%; Pred. No. 0.034;
Matches 72; Conservative 23; Mismatches 111; Indels 50; Gaps 10;

```
Qy    128  ALSAAFQAPPCAGAMAVVGGALAYLVVKTLINATQLLKLLAKL-AELVAAAIADIISDVA  186
                |||::    ||  ||  |  |  |: ||  |||  ::|  ||  |||
Db     67  AQAAAFYAQF-VQALSAGGGAYA-----AAEAAAVSPLLAPINAQFVAATGRPLIGNGA  119

Qy    187  DIIKGTLGEVWEFITNALNGLKELMDKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGL  246
                :|                        ||||||||||||||||||||:|
Db    120  NGAPGT------GANG------GPGGWLIGNGGAGGS------GAPGAGAGGNG-      155

Qy    247  SQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSFGGLPSLAQVHAASTRQAL  306
                |||||||||||||||||||||||:|||
Db    156  -------GAGGLFGSGG-------AGGASTDVAGGAGGAGGAGG--NAGMLFGAAGVGGV  199

Qy    307  RPRADGPVGAAAEQVGGQSQLVSAQSGQMGGP                -----PSSG  350
                |||||||||||||||||| :|||
Db    200  GGFSNGGATGGAGGAGGAGGLFGAGRERGSGGSGNLTGGAGGAGGNAGTLAT--GDGDGAG  257

Qy    367  GTEDAERAPVEADAGG  382
                |:||
Db    258  GTGGASRSGGFGGAGG  273
```

RESULT 8
G70917
hypothetical glycine-rich protein Rv1452c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 07-Mar-2003
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: G70917
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-741 <COL>
A;Cross-references: GB:Z95844; GB:AL123456; NID:g3250713; PIDN:CAB09269.1; PID:g2131044
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1452c
C;Superfamily: elastin related uncharacterized glycine-rich protein, PE motif containing

Query Match   7.7%; Score 151; DB 2; Length 741;
Best Local Similarity 26.6%; Pred. No. 0.035;
Matches 80; Conservative 25; Mismatches 132; Indels 64; Gaps 13;

```
Qy    110  EFVRPVAVDLTYI-PVVGHALSAAFQAPPCAGAMAVV---GGALAYLVVKTLIN------  159
                 |  |  ||  ||||  |  ||  || :|
Db      9  ETVAAAASDVARIGSSIGVANSAA-----AGSTTSVLAAGADEVSAAIATLFGSHAREY   62

Qy    160  --ATQLLKLLAKLLAELVAAAIADIISDVADIIKGTLGEVWEFITNALN------GLKE  209
                |  |  :  |  :|||  |||  ||
Db     63  QAISTQVAAFHDRFAQTLSAAVGSYVSAEA-TNAAPLATLEHNVLNALNAPTQALLGRPL  121

Qy    210  LWDKL-----TGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGL  264
                |  :|||||||||||||||||||||||||||
Db    122  IGDGAAGAPGTGQAGGAGGILWGNGGAGGSGAPGQVGGAGG--AAGLFGTGAGGAGGA  178

Qy    265  AHADSLASSASLPALAGIGGGSFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQ  324
                |||  |||  |||||||||||||||||||
Db    179  GAAGGAGGSGGW--LLGNGGVGGAGG-------QSL-------LGGATGGAGGN  216

Qy    325  SQLVSAQGSGQMGGPVGMGGMHPSSGAS-KGTTTKKYSEGAAAGTEDAERAPVEADAGG  382
                :|  ||  |||||
Db    217  AGLFGVGGTGGPGGPGGVGGTGGAGGLGGTLYGAGGHGGAGGP-----GPIGGVGGH  271

Qy    383  G  383
           |
Db    383  G  383
```

Db   272 G 272

RESULT 9
A70934
hypothetical glycine-rich protein Rv0578c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: A70934
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70934
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1306 <COL>
A;Cross-references: GB:AL021942; GB:AL123456; NID:g3242298; PIDN:CAA17449.1; PID:g2909637
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0578c
C;Superfamily: collagen alpha 1(IV) chain

Query Match   7.6%; Score 150; DB 2; Length 1306;
Best Local Similarity 28.3%; Pred. No. 0.08;
Matches 60; Conservative 10; Mismatches 82; Indels 60; Gaps 8;

```
Qy    216  GWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSAS  275
                ||||
Db    185  GWLYGNGGAGG-----FGGA-GAVGGNGGAGGTAGLFGVGAGGAGGNGIAGVTGTSAS  237

Qy    276  LP------ALAGIGGGSFGGL-----------------PSLAQVHAASTRQALRPRA  310
                |:
Db    238  TPGGSGTAGGAGGIGNGGAGGAGGVLMGNGGNGAGGEGGPGGAGGAGASGAHATNLGA  297

Qy    311  DGPVGAAAEQVGGQSQLVSAQSGQMGGP-----VGMGGMH------------PSSG  350
                |  |:|||
Db    298  DG------QAGGNGNGGAGGTGGVGGPGGHGLLGLGGSHGAGGAGGSSGDGAGPGDG  350

Qy    351  ASKGTTTKKYSEGAAAGTEDAERAPVEADAGG  382
                :|||
Db    351  GNGATGTWGHNLG-AGGTGGNGGNPGAGGAGG  381
```

RESULT 10
T03323
gene 116 protein - Lactococcus phage bIL170
C;Species: Lactococcus phage bIL170
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 05-Nov-1999
C;Accession: T03323
R;Crutz-Le Coq, A.M.; Cesselin, B.; Commissaire, J.; Anba, J.; Kyriakidis, S.; Chopin, M.
submitted to the EMBL Data Library, June 1997
A;Description: Sequence and organization of the lactococcal isometric bIL170 phage genome
A;Reference number: Z14903
A;Accession: T03323
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-916 <CRU>
A;Cross-references: EMBL:AF009630; EMBL:AF009530; NID:g3282260; PIDN:AAC27195.1; PID:g3282276
C;Genetics:
A;Gene: 116

Query Match   7.6%; Score 149.5; DB 2; Length 916;
Best Local Similarity 21.8%; Pred. No. 0.057;
Matches 89; Conservative 58; Mismatches 144; Indels 117; Gaps 18;

```
Qy      1  MSRAFIIDPTISA-----IDGLYDLLGIGIPNQGGILYSSLEYFE---KALEE-------   45
                :|  |  |:|
Db    357  LSQALI--PTLEALTPVINALIDKMDDWGKGAGKAIENVKYFQDLFKQLQQNGAITQFS  414
```

```
Qy   46 -----LAAAFPG-DGWLGSAADKYAG-------KN------RNHVNFFQELADLDRQLISLIH  89
                :||:          :|: |: |                    ||
Db  415 AIWDNLKSAFGSVIGIIGNLLIKSFAGVDESTSKNSTSVENVANTISSLANKFADITKKIA 474

Qy   90 DQANAVQTTRDILEGAKKGLEFVRPVAVDL-------TYIPVV-GHALSAAFQAP        136
                                         :|:    ||:|||
Db  475 DFIGKISESKSAMDAIKVALVALAGAFVAMKVINGLIKAYETYNKIVEAGTIIQGAFNAI 534

Qy  137 FCAGAMAVVGGALAYLVVKTLINATQL--LKLLAKLAELVAAA---------IA       179
              :|:||  ||:||:|
Db  535 MAVNPFVLLGIAIAAVVAGLVYFFTQTETGKKAWASFVDFLKSAWDGIVSFFSGIGQWFA 594

Qy  180 DIISDVADIIKGT--------LGEVWEFI------TNALNGLKELWDKLTGW          217
                                   |||:|:
Db  595 DIWNGAVDGAKGIWQGLVDWFIGIVQGIQNIWNGITTFFTTLWTTVIGGIQSVWGGVTGF 654

Qy  218 VTGLF--------SRGWSNLESFFAGVPGLTG---ATSGLSQVTGLFGA            255
        :|:|:                                   :|:|:
Db  655 FSGIFDAVSSVVSTVFSAIGFASSAWNVLVSVWSAVAGFFGGIFNAVSGV--VSSVFSA  712

Qy  256 AGLASSGLAHADSLASSAS-----LPALAGIGGG---SGFGGLPSLA             295
          :|:|: :                    ||:|||
Db  713 IGSFASSAAWGVVQSIWSAVSVGFFSGIFNSVSSVVSGVFSALGGFASNA          760
```

RESULT 11
H70663
hypothetical glycine-rich protein Rv1840c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 31-Mar-2003
C;Accession: H70663
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: H70663
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-515 <COL>
A;Cross-references: GB:Z83859; GB:AL123456; NID:g3261678; PIDN:CAB06114.1; PID:g1781207
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv1840c
C;Superfamily: uncharacterized glycine-rich protein, PE motif containing

```
Query Match           7.5%;  Score 148.5;  DB 2;  Length 515;
Best Local Similarity  22.7%;  Pred. No. 0.034;
Matches  96;  Conservative  34;  Mismatches 161;  Indels 131;  Gaps  18;

Qy    4 AFIIDPTISAIDGLYDLLGIPNQGGILYSSLEYFEKALEBLAAAFPGDGWLGSAADKY    63
                                    :|:         :|: | |:|:
Db    2 SFVVAAPEVVVAAASDLAGIG---------SAI----GAANAAAVPIMGVLAAGADEV    47

Qy   64 AGKNRN---HVNFFQELADLDRQLISLIHDQANAVQTRDILEGAKKGLEFVRPVAVDL  119
                                                         |:
Db   48 SAAVADLFGAHAQAYQALS---AQAALFHEQFVHAMTAG--AGAYAGAE----AADA    95

Qy  120 TYIPVVGHALSAAFQAPFCA-GAMAVVG----GALAYLVVKTLINATQLLKLLAKLAEL  173
              :|:||                :|:|||
Db   96 AALDVLNGPFQALFGRPLIGDGANGAPGQPGGPGGLLY-------                133

Qy  174 VAAAIADIISDVADIIKGTLGEVW-----EFITNALNGLKELWDKLTGWVTGLFSRGWS  227
Db  134 -------CNGNGNGGIGQPGGAGGDAGLIGNGNG---------GIGQGPGAT         171

Qy  228 NLESFFAGVPGL-------TGATSGL---SQVTGLFGAAGLSASSGLAH           266
Db  172 GLAGAGGVGGLLFDGDGNGAGGAGGLGTGVGATCGIGPGPGGAAVGLFGHGGAGGAGGLGK 231

Qy  267 ADSLASSASLPALAGI--GGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQV-G  322
```

```
Db  232 AGFAGGAGGTGGTCGLLYGNGNGNVFSGA--------------                   
                     |||                            |:|
Db  323 GQSQLVSAQGSQ-GMGGPVGMGGMHPSSGASKGTTTKKYSEGAAAGTEDAERAPVEADAG  381
                     ||:|                             :|:|:||:||
Db  278 GDGGSVGAAPTGIGNGNGNGGNLYGDGGSGGSTLQGFSDGGTGG-----NAGMFGDGG   332

Qy  382 GG  383
Db  333 NG  334
```

RESULT 12
A70812
hypothetical glycine-rich protein Rv0833 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 07-Mar-2003
C;Accession: A70812
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70812
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-749 <COL>
A;Cross-references: GB:AL022004; GB:AL123456; NID:g3261550; PIDN:CAA17639.1; PID:g291689
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0833
C;Superfamily: elastin related uncharacterized glycine-rich protein, PE motif containing

```
Query Match           7.5%;  Score 148.5;  DB 2;  Length 749;
Best Local Similarity  31.2%;  Pred. No. 0.052;
Matches  53;  Conservative  11;  Mismatches  83;  Indels  23;  Gaps  4;

Qy  220 GLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPAL  279
                |:|:    ||:|
Db   27 GVGGAGGAGGDSAVAGVIGGAGGAGGAALLFAGGAGGAGGSGSGGAGGAGGAGGAGGAGGL  86

Qy  280 AGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGA--AAEQVGGGQSQLVSAQGSQGMG  337
              :|:|||
Db   87 FASGSSGGFGGFASTGTGGAGGT-------GGAGGLFASGGVSGTGGGAGGSGTGGVG    137

Qy  338 GPVGMGGMHPSSGA-----SKGTTTKKYSEGAAAGTEDAERAPVEADAGG           382
                ||:|                                 :|:|:||:||
Db  138 GTGGAGGLFASGGAGGAGGSGGT-------GGAGGTGGAGGLFGAGGAGG          180
```

RESULT 13
A70647
probable PPE protein - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999
C;Accession: A70647
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S
.; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: A70647
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-409 <COL>
A;Cross-references: GB:Z83867; GB:AL123456; NID:g3261695; PIDN:CAB06293.1; PID:e1299813;
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: PPE

RESULT 14
F70971
hypothetical glycine-rich protein Rv3367 - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 31-Mar-2003
C;Accession: F70971
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: F70971
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-538 <COL>
A;Cross-references: GB:AL009198; GB:AL123456; NID:g3242262; PIDN:CAA15752.1; PID:e120228
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv3367
C;Superfamily: uncharacterized glycine-rich protein, PE motif containing

```
Query Match           7.5%;    Score 148;   DB 2;    Length 588;
Best Local Similarity  24.4%;  Pred. No. 0.042;
Matches  78;  Conservative  29;  Mismatches 121;  Indels  90;  Gaps  13;

Qy 119  LTYIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAI  178
        :::: || ||| :||                 ||:|  ||: |:||: :  :|
Db   1  MSFVVAVPEALAAA------ASDVANIGSALSAANAAAAAGTT--GLLAAGADEVSAAL  51

Qy 179  ADIISDVA----------DIIKGTLG-------EVWEFITNALN--------  205
           :||| |           |:||||       ||| |||:|||
Db  52  ASLFSGHAVSYQQVAAQATALHDQFVQALTGAGGSYALTEAANVQQNLLNAINAPTQALL  111

Qy 206  GLKELWDKLTGWVTG--------LFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAG  257
         |  |:   | ||         ||||||||||||||||
Db 112  GRPLIGDGAVGTASSPDGQDGGLLFGNGAGVNS--AATPGNAGGNGNA---GLIGNGG  166

Qy 258  LSASSGLAHADSLASSASLPALAIGIGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAA  317
        |  ||| |  || ||                        |:
Db 167  TGGSGGAGAAGGAGGGGGW--LYGNGGNGIGG--------------------NAIVAGGAGNGGA  211

Qy 318  AEQVGGGQSQLVSAQSGQMGGPVGMGGMHPSSGASKGTTTKKYSEGAAA-----------  366
          |  ||:  |  :|
Db 212  ----GGAAGLWGSGGSGGQGG---NGLTGNDGVNPAPVINPALNGAAGDSNIEPQTSVL  263

Qy 367  -GTEDAERAPVEADAGGG  383
Db 264  IGTQGGDGTPGGAGVNGG  281
```

RESULT 15
D70835
hypothetical glycine-rich protein Rv0278c - Mycobacterium tuberculosis (strain H37RV)
C;Species: Mycobacterium tuberculosis
C;Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
C;Accession: D70835
R;Cole, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.
; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd, S.;
Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.
Nature 393, 537-544, 1998
A;Authors: Sqares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
A;Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome
A;Reference number: A70500; MUID:98295987; PMID:9634230
A;Accession: D70835
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-957 <COL>
A;Cross-references: GB:AL021930; GB:AL123456; NID:g3261524; PIDN:CAA17353.1; PID:g290946
A;Experimental source: strain H37Rv
C;Genetics:
A;Gene: Rv0278c
C;Superfamily: elastin

```
Query Match           7.5%;    Score 148;   DB 2;    Length 957;
Best Local Similarity  27.9%;  Pred. No. 0.074;
Matches  75;  Conservative  13;  Mismatches 108;  Indels  73;  Gaps  10;

Qy 128  ALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIADIISDVAD  187
        || | :|| :  |  || || ||
Db  64  ALSAQAQAFHAQFVQALTSGGGAY-----------------AAAEEAAVSPLLDPINE  104

Qy 188  IIKGTLGEVWEFITNALNGLKELWDKLT-----------GWVTGLFSRGWSNLESFFAGVPGLT  240
         :  |:|||  | |||
Db 105  FFLANTGR--PLIGNGANGAPG-----TGANGGDGGWLIGNGGAGGS-------GAAGVN  150

Qy 241  GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGLPSLAQVHAA  300
        |  || |      ||| |  ||  |  |  |                  || ||
Db 151  GGAGNGGAGGNGGAGGLIGNGGAGGVASS-------GIGGSGGAGGNAMLFGAGGA  203

Qy 301  STR---QALRPRADGPVGA--------AEQVGGQ-----SQLVSAQSGQMGGPVGM  342
                  |  | ||  |
Db 204  GGAGGGVALTGGAGGAGGAGGAGGNAGLLFGAAGVGGAGGVGGAGGFTNGSALGGAGGAGAGLFAT  263

Qy 343  GGMHPSSGASKGTTTKKYSEGAAAGTEDA  371
        ||:
Db 264  GGVGGSGGAG------SSGGAGGAGGA  284
```

Search completed: November 5, 2003, 17:00:58
Job time : 31 secs

The following alignment appears in the left column and corresponds to RESULT 14 (F70971):

```
Query Match           7.5%;    Score 148;   DB 2;    Length 409;
Best Local Similarity  24.2%;  Pred. No. 0.028;
Matches  94;  Conservative  36;  Mismatches 137;  Indels 122;  Gaps  15;

Qy  12  SAIDGLYDLLGIGIPNQGGI--LYSSLEYFEKALEELAAAFPGDGWLGSAADKYAGKNR  68
Db  29  AAWDGLAEELGTRAAQSFASVTAGLAGQAWQGPAALAMAAAAAPYAGWLTAAAAQSAG---  85

Qy  69  NHVNFFQELADLDRQLISLIHDQANAVQTRDIIEGAKKGLEVVRPVAVDLTYIPVVGHA  128
Db  86  ------------------AAGQARAVAS--IFEAAQAA------------TVLP----  107

Qy 129  LSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIADIISDVADI  188
Db 108  ------------AAVAANRDAFVQLV-----------MTNLFGQNAPLIIAAA--------  136

Qy 189  IKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGV-PGLTGATSGLS  247
Db 137  -EGVYEEMWAADVAAMSGYYSGASAIAAQVV-----PWASLLQRFPGLGAGATGATGGES  190

Qy 248  QVTGL-----------FGAAGLSAS-SGLAHADSLASSASL---PALAGIGGGSGFG  289
Db 191  VGTGATGGESVGTGGGESVGTGGATASGGGVGYVGSGVVASAGLAAGDPAHGSVGQGNPGG  250

Qy 290  GLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQSGQMGGPVGMGGMHPSS  349
Db 251  -----HAGVVSPGFIGAPLALAALGQMARGG--------------------T  293

Qy 350  GASKGTTTKKYSEGAAAGTEDAERAPVEA  378
Db 294  NSAPGTATE-----SARAPEPAASAPPEA  317
```

OM protein - protein search, using sw model

Copyright   GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: November 5, 2003, 16:58:57 ; Search time 19 Seconds
                                           (without alignments)
                                           970.235 Million cell updates/sec

Title: US-09-688-672A-14
Perfect score: 1969
Sequence: 1 MSRAFIIDPTISAIDGLYDL.........RAPVEADAGGQKVLVRNVV 392

Scoring table: BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :   SwissProt_41:*

     Pred. No. is the number of results predicted by chance to have a
     score greater than or equal to the score of the result being printed,
     and is derived by analysis of the total score distribution.

                                SUMMARIES

Result   %
 No.   Score Match Query Length DB   ID   Description
------ ----- ----- ------ ------ ---- ------------ -----------
   1   153   7.8   836   1   VG26_BPML5   Q05233 mycobacteri
   2   152   7.7   801   1   Y747_MYCTU   O53810 mycobacteri
   3   148.5 7.5   515   1   YI40_MYCTU   Q50594 mycobacteri
   4   148   7.4   957   1   Y278_MYCTU   P56877 mycobacteri
   5   145   7.4   914   1   WA22_MYCTU   O06794 mycobacteri
   6   142   7.2   778   1   YQ34_MYCTU   P71933 mycobacteri
   7   141.5 7.2   435   1   YU21_MYCTU   O53268 mycobacteri
   8   141   7.2   836   1   VG26_BPMD2   O64220 mycobacteri
   9   139.5 7.1   543   1   YP91_MYCTU   Q50630 mycobacteri
  10   137.5 7.0   1901   1   YZ08_MYCTU   O53553 mycobacteri
  11   132.5 6.7   463   1   YA68_MYCTU   O53416 mycobacteri
  12   132.5 6.7   864   1   ELS_RAT   Q99372 rattus norv
  13   132.5 6.7   2333   1   PGCA_CANFA   Q28843 canis famil
  14   129.5 6.6   730   1   ELS_HUMAN   P15502 homo sapien
  15   128.5 6.5   558   1   YJ83_MYCTU   Q10873 mycobacteri
  16   127   6.5   1156   1   GLH4_CAEEL   O76743 caenorhabdi
  17   127   6.4   498   1   YI18_MYCTU   Q50615 mycobacteri
  18   126.5 6.4   860   1   ELS_MOUSE   P54320 mus musculu
  19   126.5 6.4   2038   1   FSH_DROME   P13709 drosophila
  20   125.5 6.4   1758   1   CA24_CAEEL   P17140 caenorhabdi
  21   125   6.3   603   1   YD25_MYCTU   O10637 mycobacteri
  22   124.5 6.3   1763   1   CA24_ASCSU   P27393 ascaris suu
  23   124   6.3   1356   1   CA21_ONCMY   O93484 oncorhynchu
  24   124   6.3   5263   1   FBOH_BOMMO   P05790 bombyx mori
  25   123.5 6.3   747   1   SPD1_NEPCL   P19937 nephila cla
  26   123   6.2   306   1   EC40_DAUCA   Q07322 daucus caro
  27   122.5 6.2   1372   1   CA21_MOUSE   Q01149 mus musculu
  28   121.5 6.2   1691   1   CA64_HUMAN   Q14031 homo sapien
  29   120.5 6.1   1372   1   CA21_RAT   P02466 rattus norv
  30   119.5 6.1   450   1   SWP1_ENCCU   Q9XvV1 encephalito
  31   119   6.0   864   1   KLTK_HUMAN   P29376 homo sapien
  32   119   6.0   1707   1   CA24_MOUSE   P08122 mus musculu
  33   119   6.0   2124   1   PGCA_RAT   P07897 rattus norv

  34   118.5 6.0   747   1   ELS_BOVIN   P04985 bos taurus
  35   118.5 6.0   1150   1   APMU_PIG   P12021 sus scrofa
  36   118.5 6.0   1364   1   CA21_BOVIN   P02465 bos taurus
  37   117.5 6.0   460   1   MB48_MYCTU   Q933k8 mycobacteri
  38   117.5 6.0   1736   1   CA2B_HUMAN   P13942 homo sapien
  39   117   5.9   536   1   SR54_YARLI   Q99150 yarrowia li
  40   116.5 5.9   374   1   PSPD_MOUSE   P50404 mus musculu
  41   116.5 5.9   483   1   MRCO_MESAU   Q9wrb9 mesocricetu
  42   116.5 5.9   641   1   EBN1_EBV   P03211 epstein-bar
  43   116.5 5.9   1670   1   CA34_HUMAN   Q01955 homo sapien
  44   116   5.9   491   1   YK98_MYCTU   Q10707 mycobacteri
  45   116   5.9   1027   1   CAFF_RIFPA   P30754 riftia pach

                           ALIGNMENTS

RESULT 1
VG26_BPML5
ID   VG26_BPML5   STANDARD;   PRT;   836 AA.
AC   Q05233;
DT   01-FEB-1994 (Rel. 28, Created)
DT   01-FEB-1994 (Rel. 28, Last sequence update)
DT   01-FEB-1994 (Rel. 28, Last annotation update)
DE   Minor tail protein GP26.
GN   26.
OS   Mycobacteriophage L5.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;
OC   L5-like viruses.
OX   NCBI_TaxID=31757;
RN   [1]
RP   SEQUENCE FROM N.A., AND SEQUENCE OF 1-10.
RX   MEDLINE=93211282; PubMed=8459766;
RA   Hatfull G.F., Sarkis G.J.;
RT   "DNA sequence, structure and gene expression of mycobacteriophage L5:
RT   a phage system for mycobacterial genetics.";
RL   Mol. Microbiol. 7:395-405(1993).
CC   -!- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Z18946; CAA79402.1; -.
DR   PIR; S30971; S30971.
FT   INIT_MET   0   0
SQ   SEQUENCE   836 AA; 86258 MW; 52E3040AA42BAD28 CRC64;

Query Match   7.8%; Score 153; DB 1; Length 836;
Best Local Similarity   22.3%; Pred. No. 0.017;
Matches 101; Conservative 59; Mismatches 157; Indels 136; Gaps 18;

Qy   13 AIDGLYDLLGIGIPNQGGILYSSLEYFEKALELEAAAFPGDWLGSAADKYAGNRNHVN 72
               : ::::   ::   ::: ::: ::   ::: : :::   ::
Db   309 AMQGLSQTHG--------------SVINLFNRLMESGLQAM---GQLG--------GPLSTFIN 347

Qy   73 FFQE------------------------LADLDRQLISLIHDQANAVQTTRDILEGAKKG-LEFV 112
               ::   :: :   :: :|
Db   348 GFGDLFVSLMPALTSVSGLIGNVLGTLGTQLAPIVTALTPAFQTLASTLGTMLTGALQAL 407

Qy   113 RPVAVDLTYI-PVVGHALSAAFQA-------------------PFCAGAMAV 144
               : || : || : :: |::
Db   408 GPI---LTQVATLIGTTLNTALQALQPMLPSLMCSFQQISDVLVTSLAPHIPALATALGQ 464

Qy   145 VGGA---LAYLVVKTLINATQLKLLAKLAEL------------------V 174
               ::   :: || ::| |: :|:|
Db   465 VAGAVLQLAPTIISTLVPA--FVQLVPKVAELVPTIVNLVQSFANLMPVVLPLAQALVSV 522

Qy   175 AAAIADIISDVADIIKGTLGEWWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFA 234
               :: ::   :::| || : || :::   :: :| ||

```
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31    801       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
                                  RV0747.
FT   CONFLICT    188    188       G -> S (IN REF. 2).
FT   CONFLICT    225    225       R -> G (IN REF. 2).
FT   CONFLICT    227    227       R -> G (IN REF. 2).
FT   CONFLICT    295    295       K -> R (IN REF. 2).
FT   CONFLICT    300    300       S -> P (IN REF. 2).
FT   CONFLICT    338    338       T -> I (IN REF. 2).
FT   CONFLICT    377    377       A -> T (IN REF. 2).
FT   CONFLICT    577    577       T -> A (IN REF. 2).
FT   CONFLICT    580    595       MISSING (IN REF. 2).
SQ   SEQUENCE   801 AA;  65407 MW;  EA54C9BF45A00F41 CRC64;

Query Match             7.7%;  Score 152;  DB 1;  Length 801;
Best Local Similarity   28.1%;  Pred. No. 0.018;
Matches  72;  Conservative  23;  Mismatches 111;  Indels  50;  Gaps  10;

Qy   128 ALSAAFQAPFCAGAVAVVGGALAYLVVKTLINATQLLKLIAKL-AELVAAAIADIISDVA 186
             ::  : :   ::  |||  || | ::   | | ::  :   |
Db    67 AQAAAFYAQF-VQALSAGGGAYA-----AAEAAAVSPLLAPINAQFVAATGRPLIGNGA 119

Qy   187 DIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVGLTGATSGL 246
                       |:|          :::      ::
Db   120 NGAPGT--------GANG-----GPGGWLIGNGGAGGS------GAPGAGAGGNG- 155

Qy   247 SQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFPGGLPSLAQVHAASTRQAL 306
              ||  || :       : |: :         :
Db   156 ------GAGGLFGSGG------AGGASTDVAGGAGGAGGAGG-NAGMLFGAAGVGGV 199

Qy   307 RPRADGPVGRAAEQVGGQSQLVSAQSQGMGGPVGMGGMHPSSGASKGTTTKKYSEGAAA 366
                    |      ::  :        ||       :|  |
Db   200 GGFSNGGATGGAGGAGGAGGLFGAGRERGSGGSGNLTGGAGGAGGNAGTLAT--GDGGAG 257

Qy   367 GTEDAERAPVEADAGG 382
              ||    | :: ||
Db   258 GTGGASRSGGFGGAGG 273
```

```
Db   523 AGAVIQVGVSIGGALIGALANLTEIISNVIK------KVSEWVSS-FSSGAQQIAAKAA 574
              :||:          ||    |             ::|  :  ||     : |
Qy   235 GVPGL--------TGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGG 284
             : ||             :| :   |: :||        ::
Db   575 ELPGMIQSALANLMAIGLQAGKDLVQGLINGIGGMVSAAVNKAKELASSVA------GA 627

Qy   285 GSGFGGLPSLAQVHAASTRQALRPRADGPVGRAAEQVGGQSQLVSAQSQGMGGPVGMGG 344
             ||| ||:            :|:  :: |
Db   628 VKGFLGIESPSKLFTEYGQFTAEGFGNG-MEAGFKPVIERAKDLAAELSRAM----ESG 681

Qy   345 MHPSSGASKGTTTKKYSEGAAAGTEDAERAPVE 377
                          ||     |||   |
Db   682 TDP-SGILAGLDQNELKQMLAALEERKRLKVE 713
```

RESULT 2
```
ID   Y747_MYCTU   STANDARD;      PRT;   801 AA.
AC   O53810;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv0747 precursor.
GN   RV0747 OR MT0772.5 OR MTV041.21.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Allard D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation - 
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; AL021958; CAA17514.1; -.
DR   EMBL; AE006968; AAK45011.1; -.
DR   PIR; F70824; F70824.
DR   TIGR; MT0772.5; -.
DR   TubercuList; Rv0747; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Repeat; Signal; Complete proteome.
```

RESULT 3
```
ID   YI40_MYCTU   STANDARD;      PRT;   515 AA.
AC   Q50594;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv1840c.
GN   RV1840C OR MT1888 OR MTCY1A11.04 OR MTCY359.33.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Allard D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
```

RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC   SUBFAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; Z83859; CAB06114.1; -.
DR   EMBL; AE007047; AAK46159.1; -.
DR   PIR; H70663; H70663.
DR   TIGR; MT1888; -.
DR   TubercuList; Rv1840c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM   1   21   POTENTIAL.
FT   TRANSMEM   165   185   POTENTIAL.
FT   TRANSMEM   199   219   POTENTIAL.
SQ   SEQUENCE   515 AA; 43916 MW; 2EDC8B6C6D28B7F3 CRC64;

Query Match   7.5%; Score 148.5; DB 1; Length 515;
Best Local Similarity 22.7%; Pred. No. 0.019;
Matches 96; Conservative 34; Mismatches 161; Indels 131; Gaps 18;

Qy   4 AFIIDPTISAIDGLYDLLGIGIPNQGILYSSLEYFEKALEELAAAPPGDGWLGSAADKY 63
            ||:|: ||| | |:| |||||:||||
Db   2 SFVVAAPEVVVAAASDLAGIG--------SAI----GAANAAAAVPTMGVLAAGADEV 47

Qy   64 AGKNRN----HVNFFQELADLDRQLISLIHDQANAVQTRDILEGAKKGLEFVRPVAVDL 119
            : | |:| |:|| | ||:
Db   48 SAAVADLFGAHAQAIVQALS---AQAALFHEQFVHAMTAG--AGAYAGAE----AADA 95

Qy   120 TYIPVVGHALSAAFQAPFCA-GAMAVVG----GALAYLVVKTLINATQLLKLLAKLAEL 173
            :| | || |
Db   96 AALDVLNGPFQALFGRPLIGDGANGAPGQPGGGLLY----------------------- 133

Qy   174 VAAAIADIISDVADIIKGTLGEVW-----EFITNALNGLKELWDKLITGWVTGLFSRGWS 227
            |:| |||: ||| |:| :| |||||
Db   134 --------GNGGNGCNGGIQPGGAGGDAGLIGNGGNG---------------GIGGPGAT 171

Qy   228 NLESFFAGVPGL------------TGATSGL---SQVTGLFGAAGLSASSGLAH 266
            |||||||||| |:| ||| ||||| ||||| ||
Db   172 GLAGGAGGVGGLLFGDGGNGGAGGLGTGVGATGGIGGPGGAAVGLFGHGGAGGAGGLGK 231

Qy   267 ADSLASSASLPALAGI--GGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQV-G 322
            || |||| || ||| ||| ||
Db   232 AGFAGGAGGGDGGTGGTGGLLYGNGGNGNVPSGA-----ADGGAGGGDARLIGNG 277

Qy   323 GQSQIVSAQGSQ-GMGGPVGMGGNMPSSGASKGTTTKKYSEGAAAGTEDAERAPVEADAG 381
            |: | | || || ||:|
Db   278 GDGGSVGAAPTGIGNGGNGGNGGWLYGDGGSGGSTLQGFSDGGTGG----NAGMFGDGG 332

Qy   382 GG 383
           |
Db   333 NG 334

Query Match

RESULT 4
Y278_MYCTU
ID   Y278_MYCTU   STANDARD;   PRT;   957 AA.
AC   P55877;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)

DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv0278c precursor.
GN   RV0278C OR MT0291 OR MTV035.06C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC   SUBFAMILY.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; AL021930; CAA17353.1; -.
DR   EMBL; AE006936; AAK44511.1; ALT_INIT.
DR   PIR; D70835; D70835.
DR   TIGR; MT0291; -.
DR   TubercuList; Rv0278c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
KW   Hypothetical protein; Repeat; Signal; Complete proteome.
FT   SIGNAL   1   30   POTENTIAL.
FT   CHAIN   31   957   HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT   RV0278C.
FT   CONFLICT   40   40   M -> I (IN REF. 2).
FT   CONFLICT   158   158   MISSING (IN REF. 2).
FT   CONFLICT   807   807   R -> G (IN REF. 2).
SQ   SEQUENCE   957 AA; 81905 MW; 71EBABD417FBA47C CRC64;

Query Match   7.5%; Score 148; DB 1; Length 957;
Best Local Similarity 27.9%; Pred. No. 0.04;
Matches 75; Conservative 13; Mismatches 108; Indels 73; Gaps 10;

Qy   128 ALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIADIISDVAD 187
            ||:||:||| |:| |||||:||||
Db   64 ALSAQAQAPHAQFVQALTSGGAY------------------------AAAEAAAVSPLLDPINE 104

Qy   188 IIKGTLGEVWEFITNALNGLKELWDKLT------GWVTGLFSRGWSNLESFFAGVPGLT 240
            | |:||| || |||| |||||
Db   105 FFRANTGR--PLIGNGANGAPG----TGANGGDGGWLIGNGGAGGS----GAAGVN 150

Qy   241 GATSGLSQVTGLFGAAGLSASSGLAHDSLASSSALPALAGIGGGSGFGGLPSLAQVHAA 300

```
FT   CHAIN            31     914       WAG22 ANTIGEN.
SQ   SEQUENCE   914 AA;  74354 MW;  F6953C3DBE8E6AC8 CRC64;

     Query Match             7.4%;     Score 145;  DB 1;  Length 914;
     Best Local Similarity  24.6%;     Pred. No. 0.059;
     Matches   87;  Conservative  24;  Mismatches 139;  Indels 104;  Gaps  15;

Qy     76 ELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLTYIPVVG----------- 126
          |:|||       :||            ||      ||||||:||::
Db     17 DLADLGSTIAGA--NAAAANTSLLAAGADE------------ISAAIAALFGAHGRAYQAA  65

Qy    127 HALSAARFQAPFCAGAMAVVGGALA--YLVVKTLINATQLLKLLAKLAELVAAAIADIIS 183
          : :|||                         ||||||     | |   |:: :
Db     66 SAEAAAFHGRF-VQALTTGGGAYAAEAAAAVTPLLNSIN------APVLAATGRPLIG 116

Qy    184 DVADIIKGTLGEVWEFITNALNGLKELWDKLJGWTGLFSRGWSNLESFFACVPGLTGAT 243
          :|:||          |||                ||:           |||
Db    117 NGANGAPGT-----------GANGGD-------AGWLIGNGGAGGS------GAKGANGGA 153

Qy    244 SGLSQVTGLFGAAGLSASSGLAHADS-----------------------LASSASLPALAGI 282
          ||||          |||       |||                        | ||   ||
Db    154 GGPGGAAGLFGNGGAGGAGGTATANNGIGGAGGAGGSAMLFGAGGAGGAGGAATSLVGGI 213

Qy    283 GGGSGFGG-LPSLAQVHAASTRQALRPRADGPVGAAAEQ-----------VGGGQSQLVSA 330
          |||      ||                              |                ||
Db    214 GGTGGTGGNAGMLAGAAGAGGAGGSFSTAGGAGGAGGAGGLFTTGGVGGAGGQGHTGGA 273

Qy    331 QGSQGMGGPVGMGGMHPSSG-ASKGTTTKKYSEGAAAGTEDAERAPVEADAGGG 383
          |:|      ||||      |||                     ||
Db    274 GGAGGAGGLFGAGGMGGAGGFGDHGT----LGTGGAGG--------DGGGG 312

RESULT 6
YQ34_MYCTU
ID   YQ34_MYCTU   STANDARD;   PRT;   778 AA.
AC   P71933;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv2634c.
GN   RV2634C OR MT2712 OR MTCY441.04C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
```

```
Db    151 GGAGGNGGAGGNGGAGGLIGNGGAGGAGGVASS---------GIGGSGGAGGNAMLFGAGGA 203
          ||||||       |||||||||||||||                 |||  ||   |||||
Qy    301 STR---QALRPRADGPVGA----------AAEQVGGQ------SQLVSAQGSQGMGGPVGM 342
          |||       ||||||                |||               ||:
Db    204 GGAGGGVVALTGGAGGAGGNAGLLFGAAGVGGAGGFTNGGSALGGAGGAGGAGGLFAT 263

Qy    343 GGMHPSSGASKGTTTKKYSEGAAAGTEDA 371
          ||:||  ||          ||||
Db    264 GGVGGSGGAG--------SSGGAGGAGGA 284

RESULT 5
WA22_MYCTU
ID   WA22_MYCTU   STANDARD;   PRT;   914 AA.
AC   O06794;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   WAG22 antigen precursor.
GN   WAG22 OR RV1759C OR MT1807 OR MTCY28.25C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares R., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   -!- CAUTION: REF.2 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO A
CC       FRAMESHIFT IN POSITION 85.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Z95890; CAB09322.1; -.
DR   EMBL; AE007040; -; NOT_ANNOTATED_CDS.
DR   PIR; H70987; H70987.
DR   TIGR; MT1807; -.
DR   TubercuList; Rv1759c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
KW   Antigen; Repeat; Signal; Complete proteome.
FT   SIGNAL        1        30       POTENTIAL.
```

CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; Z80225; CAB02341.1; -.
DR   EMBL; AE007103; AAK47026.1; ALT_INIT.
DR   PIR; F70963; F70963.
DR   TIGR; MT2712; -.
DR   TuberculList; Rv2634c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT   51   51   V -> L (IN REF. 2).
FT   CONFLICT   63   63   Q -> H (IN REF. 2).
FT   CONFLICT   274   274   A -> T (IN REF. 2).
SQ   SEQUENCE   778 AA; 63131 MW; DAB20FE58E4999E7 CRC64;

Query Match   7.2%; Score 142; DB 1; Length 778;
Best Local Similarity 30.2%; Pred. No. 0.076;
Matches   52; Conservative 12; Mismatches 68; Indels 40; Gaps   8;

Qy   216 GWVTGLF-SRGWSNLESFPAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSA 274
                |||||| |: |||||:| |||| |||| || :| ||||| || ||||| ||||
Db   217 GGATGLIGSSGFGGAGGAAAGV-GTTGGVGGSGVGGVFGNGGFGGAGGLGAAGGVGGAA 275

Qy   275 SLPALAGIGGGSGFGGLPSLAQVHAASTRQALRPRAD-----------------GPVGAAA 318
                :| | ||||| |||| | :| || | | | || : ||||||
Db   276 S---YFGTGGGGVGGDGA--------------PGGDGGAGPLLIGNGGVGGLGGAGAAG 318

Qy   319 EQVGGGQSLVSAQSQGMGGPV--GMGGMHPSSGASKGTTTKKYSEGAAAAG 367
                ||| | ||||| | || |||| : |||| |:| |:|| ||||| ||||
Db   319 GNGGAGGMLLGDGGAGGQGGPAVAGVLGGM-PGAGGN-GGNANWFGSGGAGG 368

RESULT 7
YU21_MYCTU
ID   YU21_MYCTU   STANDARD;   PRT;   435 AA.
AC   O53368; O53269;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PPE-family protein Rv3021c/Rv3022c.
GN   RV3021C/RV3022C OR MT3105 OR MTVO12.35C/MTVO12.36C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton J., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PPE FAMILY.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to a
CC   frameshift in position 82.
CC   ---
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---
DR   EMBL; AL021287; CAA16106.1; ALT_FRAME.
DR   EMBL; AL021287; CAA16107.1; ALT_FRAME.
DR   EMBL; AE007129; AAK47435.1; -.
DR   TIGR; MT3106; -.
DR   TuberculList; Rv3021c; -.
DR   TuberculList; Rv3022c; -.
DR   InterPro; IPR000030; Microbac_PPE.
DR   Pfam; PF00823; PPE; 1.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT   299   299   G -> A (IN REF. 2).
FT   CONFLICT   317   320   LAGV -> VTGL (IN REF. 2).
FT   CONFLICT   326   326   L -> V (IN REF. 2).
SQ   SEQUENCE   435 AA; 42876 MW; 3B15743EAA8484A CRC64;

Query Match   7.2%; Score 141.5; DB 1; Length 435;
Best Local Similarity 25.3%; Pred. No. 0.043;
Matches   97; Conservative 47; Mismatches 157; Indels 83; Gaps 18;

Qy   35 SLEYFEKA--LEELAAAFPGDGWLGSAADKYAGKNRNHVNFFQELADLDRQLISLIHDQA 92
                :|| | ||: || ||| |||| |::||||:|||||:| |::|:||||::|:: |||
Db   38 SAEYAAVAQELSVVVAAVGAGVWQGPSAELFVAAYVPYVAWLVC-ASADSAAAAGEHEAA 96

Qy   93 N-----AVQTTRDILEGAKKGLEFVRPVAVDLTYIPVVGHALSAA---------FCQAPFCAGA 141
                : |: |||| :||| |:| |::||| :||||||::| : || ||:|
Db   97 AAGVCALAEMPTLPELAANHLTHAVLVATNFFGINTIPIALNEADYVRMWVQAATVMSA 156

Qy   142 M-AVVGGAL------AYLVVK---TLINATQLLKL------LAKLAELVAAAIADI 181
                | | |||| | : || | ||| | || |::|| |:|:
Db   157 YEAVVGAAIVATPHTGPAVIVKPGANEASNAVAAATTPFPFGELAKFLEMAAQA 212

Qy   182 ISDVADIIKGTLGEVW---------EFITNALN------GLKELWDKLTGWVTGLF------ 222
                |:| | ||: ||| | ||: || | |: ||: || |||:
Db   213 FTEVGELIMKS-AEAWAVGFVELITGLVNFEPWLVLTGMIDMFFATVGFALGVFVLVPLL 271

Qy   223 --------SRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLA 271
                         :: :|| |||||| || |:||||||:|||:|||::|: |:|::::|
Db   272 EFAVVLEIAILSIGWI-ISNIFGAIPVLGGPLLGALAAAVPGVAGLAGVAGLAALPAVG 330

Qy   272 SSASLPA-----LAGIGGG--GGFGGLPSLAQVHAASTRQAL-RPRADGPVG----AAA 318
                |||:| || ||| |||| |:: :|:|| :||::|| ||||||::
Db   331 AAAGAPAALVGSVAPVSGGVVSPQARLVSAVEPAPASTSVSVLASDRGAGALGFVGTAGK 390

Qy   319 EQVGGGQSLVSAQSQGMGGPVGM 342
                |||| | |||||:||| || |
Db   391 ESVGQPAGLTVLADEFGDGAPVPM 414

RESULT 8
VG26_BPMD2
ID   VG26_BPMD2   STANDARD;   PRT;   836 AA.
AC   O64220;
DT   15-DEC-1998 (Rel. 37, Created)
DT   15-DEC-1998 (Rel. 37, Last sequence update)
DT   15-DEC-1998 (Rel. 37, Last annotation update)
DE   Minor tail protein GP26.

```
GN   26.
OS   Mycobacteriophage D29.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX   NCBI_TaxID=28369;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98300335; PubMed=9635706;
RA   Ford M.E., Sarkis G.J., Belanger A.E., Hendrix R.W., Hatfull G.F.;
RT   "Genome structure of mycobacteriophage D29: implications for phage
RT   evolution.";
RL   J. Mol. Biol. 279:143-164(1998).
```

```
DR   EMBL; AF022214; AAC18467.1; -.
DR   PIR; H72802; H72802.
FT   INIT_MET       0      0       BY SIMILARITY.
SQ   SEQUENCE   836 AA;  86572 MW;  528D12BD0FCB92E6 CRC64;

Query Match             7.2%;  Score 141; DB 1; Length 836;
Best Local Similarity  22.3%;  Pred. No. 0.095;
Matches   83;  Conservative  57;  Mismatches 137;  Indels  96;  Gaps  13;

Qy   77  LADLDRQLISLIHDQANAVQTTRD----ILEGAKKGLEFVRPVAVDLTYIPVVGHALSAA 132
Db  370  LGTLGTQLAPIITALTPAFTLADTLGTMLTGALQALGPVLTVVAE----TLGTALTTA 424

Qy  133  FQA--------------------------------------PFCAGAMAVVGGA--LAYLVVKTLINA-- 160
Db  425  LQAIQPMLPTLVDSFKQLSETLVTSLGPYLPQIGEAFGQIVGAVIQLAPTIISSLIPAFQ 484

Qy  161  ------------------TQLKLLAKLAELV---AAAIADIISDVADIKGTLGEVW 197
Db  485  TLIPAIAQLAPSLVQIVQAFTKLMPVIVPVVPVQIVINLAAAVVQAGASIASFLIGGISRLV 544

Qy  198  EFITNALNGLKELWDKLITGWVTGLFSRGWSNLESFFAGVPGLITGA---------TSGLS 247
Db  545  GVLADCVGAVAE------WV-GSWSSGVQQVSDFVGQLPGKIKSWFPDAGSWLIEAGKN 596

Qy  248  QVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGSGFGGLPSLAQVHAASTRQALR 307
Db  597  VVQGLINGIGSMISSAVSKAKELASSVK------NAVTGFLGIHSPSRVFAEIGQFTAE 649

Qy  308  PRADGPVGAAEQVGGQSQGQLVSAQGSQGMGGMHPSSGASKGTTTKKYSEGAAAG 367
Db  650  GFGNG-FEEGFQPVIEKAKALAAELSQAM-----ESGVDP-SGILAGISTKELKQYSAAL 702

Qy  368  TEDAERAPVEADA 380
Db  703  EQERKRIQVEKNA 715

RESULT 9
YP91_MYCTU
ID   YP91_MYCTU     STANDARD;      PRT;   543 AA.
AC   Q50630;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv2591.
GN   RV2591 OR MT2668.1 OR MTCY227.10C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
```

```
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
```

```
DR   EMBL; Z77724; CAB01283.1; -.
DR   EMBL; AE007100; AAK46982.1; ALT_INIT.
DR   PIR; F70726; F70726.
DR   TIGR; MT2668.1; -.
DR   TubercuList; Rv2591; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   543 AA;  46287 MW;  59730339E5D2DF59 CRC64;

Query Match             7.1%;  Score 139.5; DB 1; Length 543;
Best Local Similarity  24.1%;  Pred. No. 0.074;
Matches   77;  Conservative  34;  Mismatches 149;  Indels  59;  Gaps  11;

Qy   94  AVQTTRDILEGAKKGLEFVRPVAVDLTYIPVVGHALSAAFQAPFCAGAMAVVGGALAYLV 153
Db   33  AASTTSVLAAGADEVSQAIARLFSDYATHYQSLNAQAAAFHHSF-VQTLNAAGG--AYSS  89

Qy  154  VKTLINATQLLK--LLAKLAELVAAAIADISDVADIIKGTLGEVWEFITNALNGLKELW 211
Db   90  AEAANASAQALEQNLLA----------VINAPAQALFGR-----PLIGNGANGTAASP 132

Qy  212  DKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSL- 270
Db  133  NGGDGGI--LYGNGGNGFSQTTAGVAGGAGGLIGNGGNGGAGGAGGAGGAGGAGGWL 190

Qy  271  -----ASSASLPA------------LAGIGGSGFGGLPSL----AQVHAASTRQ 304
Db  191  LGNGGAGGPGGPTDVPAGTGGAGGDAPLIGWGNGGPGGPGGFAAFGNGGAGGNGGASGS 250

Qy  305  ALRPRADGPVGAAAEQVGGQSQLVSAGSQGMGGMHPSSGASKGTTTKKYSEGA 364
Db  251  LFGVGGAGGVGGSSEDVGG----TGGAGGARGRGLFLGLGG----DGGAGGTSNNNGGDGG 302

Qy  365  AAGTEDAERAPVEADAGGG 383
Db  303  AGGTAGGRLFSLGGDGGNG 321
```

RESULT 10

```
ID   YZ08_MYCTU     STANDARD;      PRT;   1901 AA.
AC   O53553;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv3508 precursor.
GN   RV3508 OR MTV023.15.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL022022; CAA17745.1; -.
DR   PIR; F70806; F70806.
DR   TubercuList; Rv3508; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Repeat; Signal; Complete proteome.
FT   SIGNAL        1     30       POTENTIAL.
FT   CHAIN        31   1901       HYPOTHETICAL PE-PGRS FAMILY PROTEIN
FT                                RV3508.
SQ   SEQUENCE   1901 AA;  147627 MW;  C7B1923D5D0146CD CRC64;

Query Match             7.0%;  Score 137.5;  DB 1;  Length 1901;
Best Local Similarity  26.0%;  Pred. No. 0.39;
Matches   56;  Conservative   16;  Mismatches   96;  Indels   47;  Gaps    5;

Qy    212 DKLIGWVTGLFSRGWSNLESFFAGVPGLITGATSGLSQVTGLFGAAGLSASSGLAHADSL- 270
          :: :      ::||||:|
Db    341 DGVLGGVGGTGGKGG--VGGVAGLGGAGGAAGGLFSAGGAAGAVGVGGTGGQGGGAGAG   397

Qy    271 ASSASLPALAGIGGGSGF----------------------GGLPSLAQVHAAST       302
          ||:|  || ||||||:||                      ::|| :|:|| :|
Db    398 AAGADAPASTGLTGGTGFGGAGGAGGVGGQGGNAIAGGINGSGGAGGTGGQGGAGGMGGSGA 457

Qy  , 303 RQALRPRADGPVG------------AAAEQVGGGQSQLVSAQGSQGMGGPVGMGGVHPSSGA 351
          | :| |:| |:|              :::|:|| |:  ||  :| :||||:||:||:| ||
Db    458 DNASGIGADGGAGGDGAGGAGGTGGNAGGAGGAGGAAGTGGTGGVVGAAGKAGIGGTGGQGG----AGG 513

Qy    352 SKGTTTKKYSEGAAAGTEDAEARAPVEADAGGGQKV 386
          ::|    :||||| :|||:  | | |||||||| |
Db    514 AGSAGTDATATGATGGTGFGGAGGAGGAGGNTGV 548
```

RESULT 11

```
ID   YA68_MYCTU     STANDARD;      PRT;    463 AA.
AC   O53416;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv1068c.
GN   RV1068C OR MT1097 OR MTV017.21C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AL021897; CAA17184.1; -.
DR   EMBL; AE006991; AAK45353.1; ALT_INIT.
DR   PIR; B70893; B70893.
DR   TIGR; MT1097; -.
DR   TubercuList; Rv1068c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT    218    218       T -> S (IN REF. 2).
FT   CONFLICT    235    235       G -> GGGAGIGGADGTKGGDAGAGGAGGAGGWIHGHGGV
FT                                GGDGGTGGQGGGDVQGEPGDTGAAGGAGG (IN REF.
FT                                2).
SQ   SEQUENCE    463 AA;   39305 MW;  CF5696A7E593952 CRC64;

Query Match             6.7%;  Score 132.5;  DB 1;  Length 463;
Best Local Similarity  23.5%;  Pred. No. 0.17;
Matches   72;  Conservative   31;  Mismatches  117;  Indels   87;  Gaps   10;

Qy    119 LTYIPVVGHALSAAFQAPFCAGAMAVVGGALAVLVVKTLINATQLLKLLAKLAELVAAAI 178
          : ::   | ::|:::          :  ::| | :        :: |:
Db      1 MSYMIAVPDMLSSA-----------AGDLASIGSSIN----ASTRAAAAATTRLLPAAADEVSAHI 51
```

```
Qy   179  ADIIS------------------------------DVADIIKGTLGEVWEFITNAL 204
                                              :|   ||:|:|||||||||
Db    52  AALFSGHGEGYQAIARQMAAFHDQFTLALTSSAGAYASAEATNVEQQVLGLI-NAPTQAL 110

Qy   205  NGLKELWDKLTGWVTG-------LFSRGWSNLESFPAGVPGLTGATSGLSQVTGLFGAA 256
                                    :|:    |||:|||:|||
Db   111  LGRPLIGNGADGTAANPNGGAGGLLYGNGNGFSQTTAGLITGGTGGSAGLIGNGGNGGAG 170

Qy   257  GISASSGLAHADSLASSASLPALAIGIGGSGFGLGLPSLAQVHAASTRQALRPRADGPVGA 316
                           :||:|||  ||||:|     ||  |  :|
Db   171  GAGANGG-------AGGNGGWIYGSGGNGGAGGAGPAGAIGA--------PGVAGGAGG 214

Qy   317  AAEQVGGGQSOLVSAAQGSQCMGGPVGMGGMHPSSGASKGTTTKKYSEGAAAGTEDAERAPV 376
                                :     |||          ||:|
Db   215  A----GGTAGLFGNGGAGGAGGAGAGAGG----RGGDGGSAGWLSGNGGDAGT-------- 258

Qy   377  EADAGGG 383
          ||:
Db   259  ---GGGG 262


RESULT 12
ELS_RAT
ID   ELS_RAT        STANDARD;      PRT;    864 AA.
AC   Q99372;
DT   01-OCT-1995 (Rel. 34, Created)
DT   01-OCT-1995 (Rel. 34, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Elastin precursor (Tropoelastin) (Fragment).
GN   ELN.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=91104868; PubMed=1702999;
RA   Pierce R.A., Deak S.B., Stolle C.A., Boyd C.D.;
RT   "Heterogeneity of rat tropoelastin mRNA revealed by cDNA cloning.";
RL   Biochemistry 29:9677-9683(1990).
RN   [2]
RP   SEQUENCE OF 781-864 FROM N.A.
RX   MEDLINE=88330868; PubMed=2971041;
RA   Deak S.B., Pierce R.A., Belsky S.A., Riley D.J., Boyd C.D.;
RT   "Rat tropoelastin is synthesized from a 3.5-kilobase mRNA.";
RL   J. Biol. Chem. 263:13504-13507(1988).
RN   [3]
RP   SEQUENCE OF 264-533 AND 558-864 FROM N.A., AND ALTERNATIVE SPLICING.
RX   MEDLINE=92241859; PubMed=1572637;
RA   Pierce R.A., Alatawi A., Deak S.B., Boyd C.D.;
RT   "Elements of the rat tropoelastin gene associated with alternative
RT   splicing.";
RL   Genomics 12:651-658(1992).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=8;
CC         Comment=Experimental confirmation may be lacking for some
CC         isoforms;
CC       Name=1;
CC         IsoId=Q99372-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q99372-2; Sequence=VSP_004244;
CC       Name=3;
CC         IsoId=Q99372-3; Sequence=VSP_004245;
CC       Name=4;
CC         IsoId=Q99372-4; Sequence=VSP_004246;
CC       Name=5;
CC         IsoId=Q99372-5; Sequence=VSP_004244, VSP_004245;
CC       Name=6;
CC         IsoId=Q99372-6; Sequence=VSP_004245, VSP_004246;
CC       Name=7;
CC         IsoId=Q99372-7; Sequence=VSP_004244, VSP_004246;
CC       Name=8;
CC         IsoId=Q99372-8; Sequence=VSP_004244, VSP_004245, VSP_004246;
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; M60647; AAA42269.1; -.
DR   EMBL; J04035; AAA42268.1; -.
DR   EMBL; M86372; AAA42271.1; -.
DR   EMBL; M86355; AAA42271.1; JOINED.
DR   EMBL; M86363; AAA42271.1; JOINED.
DR   EMBL; M86364; AAA42271.1; JOINED.
DR   EMBL; M86366; AAA42271.1; JOINED.
DR   EMBL; M86371; AAA42271.1; JOINED.
DR   EMBL; M86376; AAA42272.1; -.
DR   EMBL; M86373; AAA42272.1; JOINED.
DR   EMBL; M86375; AAA42272.1; JOINED.
DR   PIR; A36106; EART.
DR   InterPro; IPR003979; tropoelastin.
DR   PRINTS; PR01500; TROPOELASTIN.
KW   Structural protein; Connective tissue; Repeat; Signal;
KW   Alternative splicing.
FT   NON_TER       1      1
FT   SIGNAL        1     21                  BY SIMILARITY.
FT   CHAIN        22    864                  ELASTIN.
FT   DISULFID    854    859                  BY SIMILARITY.
FT   VARSPLIC    263    307                  Missing (in isoform 2, isoform 5, isoform
FT                                           7 and isoform 8).
FT                                           /FTId=VSP_004244.
FT   VARSPLIC    308    308                  Missing (in isoform 3, isoform 5, isoform
FT                                           6 and isoform 8).
FT                                           /FTId=VSP_004245.
FT   VARSPLIC    809    823                  Missing (in isoform 4, isoform 6, isoform
FT                                           7 and isoform 8).
FT                                           /FTId=VSP_004246.
SQ   SEQUENCE   864 AA;  72786 MW;  456894BB09E79FD4 CRC64;

Query Match             6.7%;  Score 132.5;  DB 1;  Length 864;
Best Local Similarity  23.1%;  Pred. No. 0.34;
Matches   94;  Conservative  35;  Mismatches 135;  Indels 143;  Gaps  18;

Qy     9  PTISAIDGLYDLLGI-GIPNQQGIYSSLEYFEKALEELAAAFPGDGWLG------------ 57
             ||                                                    |||||||||
Db   384  PGVGGIPGVGGIPGVGG-------------------------------PGIGGPGIVGGPGAVSP 423

Qy    58  SAADKYAGKRNHVNFFQELADLDRQLISLIHDQANAVQTRDILEGAKKGLEFVRPVAV 117
             :|||||                                       || :|
Db   424  AAAAKAAAKAAKY-----------------------------GARGGVG--------- 443

Qy   118  DLTYIPVVGHALSAAFQAPFCAGAMAVVGALAYLVVKTLINATQLLKLAKLAELVAAA 177
                       |||                                |:|||
Db   444  ----IPTYG--VGAGGPGVGVGAGAGLGGA--------------------SQAAAAA 475

Qy   178  IADIISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLPSRGWSNLESFPAGVP 237
                       :|||||||:|:                      |:||:
Db   476  AA---AKAAKYGAGAGTLGGLVPGAVPG------ALPGAVPGALP--------GAVP 516

Qy   238  G-LTGATSGLSQVTGLEGAAGLSASSGLAHADSLASSASL--PALAGIGGSGFCGLP- 292
             | ||||                              |:||:|: :||||
Db   517  GALPGAVPGVPGTGGVPGVGAGTPAAAAAAAAKAAAKAGQYGLGPGVGPGVGVGGLPG 576

Qy   293  --SLAQVHAAST--RQALRPRADGPVGAAAEQVGGQSQLVSAQGSQ------------GMGGP-V 340
             :: :|||| |                                          ::| :
```

```
Db    577 GVGPGGVTGIGTGPGTGLVP---GDIGGAGTPAAAKSAAKAAAKAQYRAAAGLGAGVPGL 633

Qy    341 GMGGMPSSGA-----SKGTTTKKYSEGAAAGTEDAERAPVEADAGG 382
          |:|           ||:|           ||:|    |:|
Db    634 GVGAGVPGFGAGAGFGAGAGVPGFGAGAVPGSLAASKAAKYGAAGG 680
```

```
RESULT 13
PGCA_CANFA
ID   PGCA_CANFA      STANDARD;      PRT;   2333 AA.
AC   Q28343; Q28310;
DT   01-NOV-1997 (Rel. 35, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Aggrecan core protein precursor (Cartilage-specific proteoglycan core
DE   protein) (CSPCP).
GN   AGC1.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Glant T.T., Adams M.E., Kwok S.F., Huang D., Fueloep C.;
RT   "Complete coding sequence and deduced amino acid sequence of aggrecan
RT   of canine cartilage.";
RL   Submitted (APR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE OF 774-833 FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=93128522; PubMed=7827755;
RA   Barry F.P., Neame P.J., Sasse J., Pearson D.;
RT   "Length variation in the keratan sulfate domain of mammalian
RT   aggrecan.";
RL   Matrix Biol. 14:323-328(1994).
RN   [3]
RP   SEQUENCE OF 1830-2333 FROM N.A.
RA   Adams M.E., Kwok S.F., Huang D., Glant T.T., Fullop C.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 2082-2118 FROM N.A.
RC   TISSUE=Cartilage;
RX   MEDLINE=93352525; PubMed=8349621;
RA   Fueloep C., Walcz E., Valyon M., Glant T.T.;
RT   "Expression of alternatively spliced epidermal growth factor-like
RT   domains in aggrecans of different species. Evidence for a novel
RT   module.";
RL   J. Biol. Chem. 268:17377-17383(1993).
CC   -!- FUNCTION: THIS PROTEOGLYCAN IS A MAJOR COMPONENT OF EXTRACELLULAR
CC       MATRIX OF CARTILAGENOUS TISSUES. A MAJOR FUNCTION OF THIS PROTEIN
CC       IS TO RESIST COMPRESSION IN CARTILAGE. IT BINDS AVIDLY TO
CC       HYALURONIC ACID VIA AN AMINO-TERMINAL GLOBULAR REGION. MAY PLAY A
CC       REGULATORY ROLE IN THE MATRIX ASSEMBLY OF THE CARTILAGE.
CC   -!- SUBCELLULAR LOCATION: Secreted; extracellular matrix (By
CC       similarity).
CC   -!- DOMAIN: TWO GLOBULAR DOMAINS, G1 AND G2, COMPRISE THE AMINO
CC       TERMINUS OF THE PROTEOGLYCAN, WHILE ANOTHER GLOBULAR REGION, G3,
CC       MAKES UP THE COOH TERMINUS. G1 CONTAINS LINK DOMAINS AND THUS
CC       CONSISTS OF THREE DISULFIDE-BONDED LOOP STRUCTURES DESIGNATED AS
CC       THE A, B, B' MOTIFS. G2 IS SIMILAR TO G1. THE KERATAN SULFATE (KS)
CC       AND THE CHONDROITIN SULFATE (CS) ATTACHMENT DOMAINS LIE BETWEEN G2
CC       AND G3.
CC   -!- PTM: CONTAINS MOSTLY CHONDROITIN SULFATE, BUT ALSO KERATAN SULFATE
CC       CHAINS, N-LINKED AND O-LINKED OLIGOSACCHARIDES (BY SIMILARITY).
CC   -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC   -!- SIMILARITY: Contains 4 link domains.
CC   -!- SIMILARITY: Contains 1 C-type lectin family domain.
CC   -!- SIMILARITY: Contains 1 Sushi (SCR) domain.
CC   -!- SIMILARITY: Contains 1 EGF-like domain.
CC   -!- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
DR   EMBL; U65989; AAB06238.2; -.
DR   EMBL; S74662; AAC60527.1; -.
DR   EMBL; L07054; -; NOT_ANNOTATED_CDS.
DR   PIR; I46998; I46998.
DR   HSSP; P08709; 1BF9.
DR   InterPro; IPR000152; Asx_hydroxyl.
DR   InterPro; IPR000742; EGF_2.
DR   InterPro; IPR001881; EGF_Ca.
DR   InterPro; IPR006209; EGF_like.
DR   InterPro; IPR007110; Ig-like.
DR   InterPro; IPR003006; Ig_MHC.
DR   InterPro; IPR001304; Lectin_C.
DR   InterPro; IPR000538; Link.
DR   InterPro; IPR003324; SGXXSG.
DR   InterPro; IPR000436; Sushi_SCR_CCP.
DR   Pfam; PF00008; EGF; 1.
DR   Pfam; PF00047; ig; 1.
DR   Pfam; PF00059; lectin_C; 1.
DR   Pfam; PF02339; SGXXSG; 66.
DR   Pfam; PF00084; sushi; 1.
DR   Pfam; PF00193; Xlink; 4.
DR   PRINTS; PR01265; LINKMODULE.
DR   PRINTS; PR00356; ANTIFREEZEII.
DR   ProDom; PD000918; Link; 4.
DR   SMART; SM00003; CCP; 1.
DR   SMART; SM00034; CLECT; 1.
DR   SMART; SM00179; EGF_CA; 1.
DR   SMART; SM00445; LINK; 4.
DR   PROSITE; PS00010; ASX_HYDROXYL; 1.
DR   PROSITE; PS00615; C_TYPE_LECTIN_1; 1.
DR   PROSITE; PS50041; C_TYPE_LECTIN_2; 1.
DR   PROSITE; PS00022; EGF_1; 1.
DR   PROSITE; PS01187; EGF_CA; 1.
DR   PROSITE; PS50035; IG_LIKE; 1.
DR   PROSITE; PS00290; IG_MHC; 1.
DR   PROSITE; PS01241; LINK; 4.
KW   Glycoprotein; Cartilage; Proteoglycan; Lectin; Signal; Sushi;
KW   EGF-like domain; Repeat; Immunoglobulin domain.
FT   SIGNAL        1     16       POTENTIAL.
FT   CHAIN        17   2333       AGGRECAN CORE PROTEIN.
FT   DOMAIN       34    147       IG-LIKE V-TYPE.
FT   DOMAIN      170    247       LINK 1.
FT   DOMAIN      268    349       LINK 2.
FT   DOMAIN      513    590       LINK 3.
FT   DOMAIN      611    692       LINK 4.
FT   DOMAIN     2081   2117       EGF-LIKE; CALCIUM-BINDING (POTENTIAL).
FT   DOMAIN     2130   2245       C-TYPE LECTIN.
FT   DOMAIN     2249   2307       SUSHI.
FT   DOMAIN       48    140       G1-A.
FT   DOMAIN      152    247       G1-B.
FT   DOMAIN      253    349       G1-B'.
FT   DOMAIN      495    589       G2-B.
FT   DOMAIN      596    691       G2-B'.
FT   DOMAIN      694    816       KS.
FT   DOMAIN      819   1394       CS-1.
FT   DOMAIN     1395   2079       CS-2.
FT   DOMAIN     2080   2333       G3.
FT   DISULFID     51    133       BY SIMILARITY.
FT   DISULFID    175    246       BY SIMILARITY.
FT   DISULFID    199    220       BY SIMILARITY.
FT   DISULFID    273    348       BY SIMILARITY.
FT   DISULFID    297    318       BY SIMILARITY.
FT   DISULFID    518    589       BY SIMILARITY.
FT   DISULFID    542    563       BY SIMILARITY.
FT   DISULFID    616    691       BY SIMILARITY.
FT   DISULFID    640    661       BY SIMILARITY.
FT   DISULFID   2085   2096       BY SIMILARITY.
```

```
FT   DISULFID     2090   2105       BY SIMILARITY.
FT   DISULFID     2107   2116       BY SIMILARITY.
FT   DISULFID     2123   2134       BY SIMILARITY.
FT   DISULFID     2151   2243       BY SIMILARITY.
FT   DISULFID     2219   2235       BY SIMILARITY.
FT   DISULFID     2250   2293       BY SIMILARITY.
FT   DISULFID     2279   2306       BY SIMILARITY.
FT   CARBOHYD      126    126       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD      239    239       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD      333    333       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD      387    387       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD      444    444       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD      620    620       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD      676    676       N-LINKED (GLCNAC...)    (POTENTIAL).
FT   CARBOHYD      747    747       N-LINKED (GLCNAC...)    (POTENTIAL).
SQ   SEQUENCE   2333 AA;  240573 MW;  8B9ED78F3508BS96 CRC64;
```

Query Match   6.7%;   Score 132.5;   DB 1;   Length 2333;
Best Local Similarity 26.4%;   Pred. No. 1;
Matches   69; Conservative 35; Mismatches 82; Indels 75; Gaps 13;

```
Qy    172 ELVAAAIADI-------------ISDVADIIKGTLGEVWEFITNALNGLKELWMDKLTGWVT 219
              |: || |               |:|| :||            || |  |  :|
Db    940 EVSASGVEDLSGLPSGEGPETSGVGDLSRLPSGEGPEVSASGV---------GDLS 988

Qy    220 GLFSRGWSNLESFFAGVPGLTG--------ATSGLSQVTGL-------FGAAGLSASSG 263
              ||| |  | |  |||      |||| :|| |           |:|:  ||||||||
Db    989 GLPS-GREGLETSTSGVEDLSGLPSGEGPEASTSGVGDLSRLPSGEGPEVSASGVEDLSG 1047

Qy    264 LAHADSLASSAS-LPALAGIGGG--------SGFGGLPSLAQVHAASTRQALRPRADGP-V 314
              :||   |||||| ||| |||          |  || |            :||
Db   1048 LPSGEGLEASASGVGDLSGLPSGEGPEASASGVGDLSRL------------PSGEGPEV 1094

Qy    315 GAAAEQ-----VGGQSQLVSAQGSQGMGG-PVGMGGVHPSS--GASKGTTKKYSEGAA 365
              |  |     |  | | :||| ||  |  |||||   |    |  :||   : :||||
Db   1095 SASGVEDLSGLSSGESPEASASGVGDLSGLPSGLPSGREGLETSASGVGDLSGLPSGEGQEASA 1154

Qy    366 AGTEDAERAP---VEADAGG 382
              :|||       ||||||
Db   1155 SGVEDLSRLPSGEGPEASASG 1175
```

```
RESULT 14
ID   ELS_HUMAN    STANDARD;       PRT;    730 AA.
AC   P15502; Q14233; Q14238;
DT   01-APR-1990 (Rel. 14, Created)
DT   01-APR-1990 (Rel. 14, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Elastin precursor (Tropoelastin).
GN   ELN.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM B).
RX   MEDLINE=87289668; PubMed=3039501;
RA   Indik Z., Yeh H., Ornstein-Goldstein N., Sheppard P., Anderson N.,
RA   Rosenbloom J.C., Peltonen L., Rosenbloom J.;
RT   "Alternative splicing of human elastin mRNA indicated by sequence
RT   analysis of cloned genomic and complementary DNA.";
RL   Proc. Natl. Acad. Sci. U.S.A. 84:5680-5684(1987).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Skin fibroblast;
RX   MEDLINE=89009960; PubMed=3171221;
RA   Fazio M.J., Olsen D.R., Kauh E.A., Baldwin C.T., Indik Z.,
RA   Ornstein-Goldstein N., Yeh H., Rosenbloom J., Uitto J.;
RT   "Cloning of full-length elastin cDNAs from a human skin fibroblast
RT   recombinant cDNA library: further elucidation of alternative splicing
RT   utilizing exon-specific oligonucleotides.";
RL   J. Invest. Dermatol. 91:458-464(1988).
```

```
RN   [3]
RP   SEQUENCE OF 164-724 FROM N.A. (ISOFORM B).
RC   TISSUE=Placenta;
RX   MEDLINE=88156138; PubMed=2831431;
RA   Fazio M.J., Olsen D.R., Kuivaniemi H., Chu M.L., Davidson J.M.,
RA   Rosenbloom J., Uitto J.;
RT   "Isolation and characterization of human elastin cDNAs, and age-
RT   associated variation in elastin gene expression in cultured skin
RT   fibroblasts.";
RL   Lab. Invest. 58:270-277(1988).
RN   [4]
RP   SEQUENCE OF 603-730 FROM N.A.
RC   TISSUE=Hippocampus, and Placenta;
RX   MEDLINE=96291399; PubMed=8689688;
RA   Frangiskakis J.M., Ewart A.K., Morris C.A., Mervis C.B.,
RA   Bertrand J., Robinson B.F., Klein B.P., Ensing G.J., Everett L.A.,
RA   Green E.D., Proeschel C., Gutowski N.J., Noble M., Atkinson D.L.,
RA   Odelberg S.J., Keating M.T.;
RT   "LIM-kinase1 hemizygosity implicated in impaired visuospatial
RT   constructive cognition.";
RL   Cell 86:59-69(1996).
CC   -!- FUNCTION: MAJOR STRUCTURAL PROTEIN OF TISSUES SUCH AS AORTA AND
CC       NUCHAL LIGAMENT, WHICH MUST EXPAND RAPIDLY AND RECOVER COMPLETELY.
CC   -!- SUBUNIT: THE POLYMERIC ELASTIN CHAINS ARE CROSS-LINKED TOGETHER
CC       INTO AN EXTENSIBLE 3D NETWORK.
CC   -!- SUBCELLULAR LOCATION: EXTRACELLULAR MATRIX OF ELASTIC FIBERS.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC         Comment=Additional isoforms seem to exist;
CC       Name=1;
CC         IsoId=P15502-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=P15502-2; Sequence=VSP_004243;
CC   -!- PTM: THE CROSSLINKS ARE MADE OF DEAMINATED LYS.
CC   -!- DISEASE: Haploinsufficiency of ELN may be the cause of certain
CC       cardiovascular and musculo-skeletal abnormalities observed in
CC       Williams-Beuren syndrome (WBS), a rare developmental disorder. It
CC       is a contiguous gene deletion syndrome involving genes from
CC       chromosome band 7q11.23.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M17282; AAC98394.1; -.
DR   EMBL; M16983; AAC98394.1; JOINED.
DR   EMBL; M17265; AAC98394.1; JOINED.
DR   EMBL; M17266; AAC98394.1; JOINED.
DR   EMBL; M17267; AAC98394.1; JOINED.
DR   EMBL; M17268; AAC98394.1; JOINED.
DR   EMBL; M17270; AAC98394.1; JOINED.
DR   EMBL; M17271; AAC98394.1; JOINED.
DR   EMBL; M17272; AAC98394.1; JOINED.
DR   EMBL; M17273; AAC98394.1; JOINED.
DR   EMBL; M17275; AAC98394.1; JOINED.
DR   EMBL; M17276; AAC98394.1; JOINED.
DR   EMBL; M17277; AAC98394.1; JOINED.
DR   EMBL; M17278; AAC98394.1; JOINED.
DR   EMBL; M17279; AAC98394.1; JOINED.
DR   EMBL; M17280; AAC98394.1; JOINED.
DR   EMBL; M17281; AAC98394.1; JOINED.
DR   EMBL; M36860; AAA52382.1; -.
DR   EMBL; M24782; AAA53190.1; -.
DR   EMBL; U6229; AAB17544.1; -.
DR   EMBL; X15603; CAA33627.1; -.
DR   PIR; A32707; EAHU.
DR   HSSP; P50099; 1ZFJ.
DR   Genew; HGNC:3327; ELN.
DR   MIM; 130160; -.
```

```
DR   MIM; 194050; -.
DR   GO; GO:0005578; C:extracellular matrix; TAS.
DR   GO; GO:0005615; C:extracellular space; TAS.
DR   GO; GO:0005201; F:extracellular matrix structural constituent; TAS.
DR   GO; GO:0008283; P:cell proliferation; TAS.
DR   GO; GO:0008015; P:circulation; TAS.
DR   GO; GO:0007397; P:histogenesis and organogenesis; TAS.
DR   GO; GO:0007585; P:respiratory gaseous exchange; TAS.
DR   InterPro; IPR003979; tropoelastin.
DR   PRINTS; PR01500; TROPOELASTIN.
KW   structural protein; Connective tissue; Repeat; Signal;
KW   Williams-Beuren syndrome; Alternative splicing.
FT   SIGNAL        1     26
FT   CHAIN        27    730       ELASTIN.
FT   DISULFID    720    725       BY SIMILARITY.
FT   VARSPLIC    472    477       Missing (in isoform 2).
FT                                /FTId=VSP_004243.
SQ   SEQUENCE   730 AA;  63260 MW;  AB06D15A567AE46 CRC64;

Query Match            6.6%;   Score 129.5;  DB 1;  Length 730;
Best Local Similarity  20.3%;  Pred. No. 0.44;  -.
Matches   86;  Conservative   36;  Mismatches  134;  Indels  167;  Gaps   15;

Qy      22 GIGIPN--QGGILYSSLEYPEKALEELAAAFPGDGWLGSAADKYAGKNRNHVNFFQELAD 79
              |:|:|                                |:||:|
Db     143 GVGLPGVPGGVL----------------------------PGARFPGVGVLPGV------- 169

Qy      80 LDRQLISLIHDQANAVQTRDILEGAKKGLEFVRPV----AVDLTYIPVVGHALSAAFQ 134
                            |:  .         :|:       |||      ||::|:
Db     170 --------------PTGAGVKPKAPGVGGAFAGIPGVGPFGGPQPGVPLGY-PIKAPKLPGGYG 218

Qy     135 APFCAGAM------AVVGGA-LAYLVVKTLINATQLLKLLAKLAELVAAAIADIISDVA 186
                            : |||          |:||:|
Db     219 LPYTTGKLPYGYGPGGVAGAAGKAGYPTGTGVGPQAAAAAAAAKFGAGAAGVLPGVG 278

Qy     187 DIIKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGVPGLTGATSGL 246
                                                      |||:||
Db     279 G---------------------------------------AGVPGVPGA----AGVPGVPGA 288

Qy     247 SQVTGLFGAAGLSASSGLAHADSLASSASLPALAG-IGGGSGFG--------GLPSL-- 294
              :  |||:|:  |     ||||                 ||:|:|
Db     289 --IPGIGGIAGVGTPAAAAAAAKAAKYGAAAGLVPGGPGFGPGVVGVPGAGVPGVGV 346

Qy     295 ----------AQVHAASTRQALRPRADGPVGAAAEQVG--------------------- 322
                        : |||         :  |:
Db     347 PGAGIPVVPGAGIPGAAVPGVVSPEAAAAAAAKAAKYGARPGVGGIPTYGVGAGGFPG 406

Qy     323 --------GQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKYSEG--AAAGTEDA 371
                      |:  |: |||         |:||:|
Db     407 FGVGVGGIPGVAGVPSVGGVPGVGVPGV-GISPAQAAAAAAAKAAKYGVGTPAAAAAKAA 465

Qy     372 ERA 374
Db     466 AKA 468


RESULT 15
YJ83_MYCTU
ID   YJ83_MYCTU     STANDARD;      PRT;   558 AA.
AC   Q10873;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical PE-PGRS family protein Rv1983.
GN   RV1983 OR MT2036 OR MTCV39.36C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=9295987; PubMed=9634230;
```

```
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- SIMILARITY: BELONGS TO THE MYCOBACTERIAL PE FAMILY. PGRS
CC       SUBFAMILY.
CC   -----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; Z74025; CAA98400.1; -.
DR   EMBL; AE007056; AAK46311.1; -.
DR   PIR; E70756; E70756.
DR   TIGR; MT2036; -.
DR   TubercuList; Rv1983; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   DOMAIN       27     53       ALA-RICH.
FT   TRANSMEM    332    352       POTENTIAL.
FT   TRANSMEM    360    380       POTENTIAL.
SQ   SEQUENCE   558 AA;  53738 MW;  17ECEB43778E021E CRC64;

Query Match            6.5%;   Score 128.5;  DB 1;  Length 558;
Best Local Similarity  25.4%;  Pred. No. 0.38;  -.
Matches   71;  Conservative   23;  Mismatches  125;  Indels  61;  Gaps   8;

Qy     110 EFVRPVAVDLTYIPVVGHALSAAFQAPFCAGAMA-----VVGGALA-----------YLVVKTL 157
             :                 :|:                                      :
Db       9 EFLTSAAADVENIGSTLRAANAAAASTTALAAAGADEVSAAVAALFARFGQEYCAVSAQ 68

Qy     158 INA--TQLLKLLAKLAELVAAAIADIISDVADIIKGTLGEVWEFITNALNGLKELWDKLT 215
             :: .                                |||
Db      69 ASAFHQQFVQTLNSASGSYAAAEATIASQLQTAQHDLLGAV-NAPTETLLGRPLIGDGAP 127

Qy     216 GWVTG-------LFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHA 267
                          |:||:|
Db     128 GTATSPNGGAGGLLYGNGGNGYSATASVGGGAGGSAGLIGNGGAGGAGGPNAPGGAG-- 185

Qy     268 DSLASSASLPALAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGGSQL 327
               |||:|:|
Db     186 --GNGGWLLGNGGIGGPGGASSIPGMS---------------------------GGAGGT 216

Qy     328 VSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKYSEGAAAG 367
               :   |:||:|     |:||
Db     217 GGAAGLLGWGANGGAGGLGDGVVDRGT-------GGAGG 249


Search completed: November  5, 2003, 17:00:21
```

us-09-688-672a-14.rsp

Thu Nov 5 11:33:28 2003

Job time : 23 secs

```
              GenCore version 5.1.6
          Copyright (c) 1993 - 2003  Compugen Ltd.

QM protein - protein search, using sw model

Run on:    November 5, 2003, 17:01:02 ; Search time 56 Seconds
                                (without alignments)
                                1806.368 Million cell updates/sec

Title:          US-09-688-672A-14
Perfect score:  1969
Sequence:       1 MSRAFIIDPTISAIDGLYDL........RAPVEADAGGGQKVLVRNVV 392

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5
Searched:       830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters:         830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :     SPTREMBL_23:*
           1:  sp_archea:*
           2:  sp_bacteria:*
           3:  sp_fungi:*
           4:  sp_human:*
           5:  sp_invertebrate:*
           6:  sp_mammal:*
           7:  sp_mhc:*
           8:  sp_organelle:*
           9:  sp_phage:*
          10:  sp_plant:*
          11:  sp_rodent:*
          12:  sp_virus:*
          13:  sp_vertebrate:*
          14:  sp_unclassified:*
          15:  sp_rvirus:*
          16:  sp_bacteriap:*
          17:  sp_archeap:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1969	100.0	392	16	O06267	O06267 mycobacteri
2	1258	63.9	394	16	Q49722	Q49722 mycobacteri
3	380.5	19.3	402	16	P96213	P96213 mycobacteri
4	168	8.5	496	16	O53631	O53631 mycobacteri
5	168	8.5	533	16	Q8VKR5	Q8vkr5 mycobacteri
6	156.5	7.9	536	16	Q8FND6	Q8FND6 corynebacte
7	155	7.9	1329	16	O06810	O06810 mycobacteri
8	152	7.7	1408	16	Q8VK17	Q8vk17 mycobacteri
9	151	7.7	738	16	Q8VK15	Q8vk15 mycobacteri
10	151	7.7	741	16	O06808	O06808 mycobacteri
11	150	7.6	1306	16	O53775	O53775 mycobacteri
12	149.5	7.6	916	9	O80116	O80116 bacteriopha
13	149	7.6	628	16	Q8VJ19	Q8vj19 mycobacteri
14	148.5	7.5	749	16	O53844	O53844 mycobacteri
15	148	7.5	409	16	P95182	P95182 mycobacteri
16	148	7.5	588	16	O50396	O50396 mycobacteri
17		7.5	148	16	1665	O53215	O53215 mycobacteri
18		7.5	147.5	16	459	Q8XQX4	Q8xqx4 ralstonia s
19		7.5	147	16	436	Q8VIX6	Q8vix6 mycobacteri
20		7.5	147	16	439	O06287	O06287 mycobacteri
21		7.4	146.5	16	837	O53684	O53684 mycobacteri
22		7.4	145.5	16	1507	Q8VJ23	Q8vj23 mycobacteri
23		7.4	145.5	16	1538	O53395	O53395 mycobacteri
24		7.4	145	16	911	Q8DXK6	Q8dxk6 streptococc
25		7.3	144.5	16	539	P71657	P71657 mycobacteri
26		7.3	144	16	1381	O53552	O53552 mycobacteri
27		7.3	144	16	1384	Q8VIZ1	Q8viz1 mycobacteri
28		7.2	142.5	16	454	Q8VJD7	Q8vjd7 mycobacteri
29		7.2	142.5	16	525	O33284	O33284 mycobacteri
30		7.2	142.5	16	667	O53415	O53415 mycobacteri
31		7.2	142.5	16	879	Q8VKD2	Q8vkd2 mycobacteri
32		7.2	142.5	16	882	O53845	O53845 mycobacteri
33		7.2	142	16	476	O06199	O06199 mycobacteri
34		7.2	141.5	16	1489	O53559	O53559 mycobacteri
35		7.2	141	16	584	O53575	O53575 mycobacteri
36		7.2	141	5	1468	Q9GUB5	Q9gub5 galleria me
37		7.1	140.5	16	853	O53439	O53439 mycobacteri
38		7.1	140	16	1463	Q9ADP6	Q9adp6 streptomyce
39		7.0	138.5	16	614	Q9LH97	Q9lh97 arabidopsis
40		7.0	138.5	16	635	Q8VK71	Q8vk71 mycobacteri
41		7.0	138.5	16	683	Q8G791	Q8g791 bifidobacte
42		7.0	137.5	16	783	O53809	O53809 mycobacteri
43		7.0	137.5	16	1715	Q8VIZ0	Q8viz0 mycobacteri
44		6.9	136	16	476	Q8VKA5	Q8vka5 mycobacteri
45		6.9	136	16	618	O06801	O06801 mycobacteri

ALIGNMENTS

```
RESULT 1
O06267
ID   O06267        PRELIMINARY;      PRT;     392 AA.
AC   O06267;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Rv3616c.
GN   RV3616C OR MT3718 OR MTCY07H7B.06.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., Mclean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
```

DR   EMBL; Z95557; CAB08950.1; -.
DR   EMBL; AE007171; AAK48077.1; -.
DR   TIGR; MT3718; -.
KW   TubercuList; Rv3616c.
KW   Hypothetical protein; Complete proteome.
FT   CONFLICT   192   192   T -> I (IN REF. 2).
SQ   SEQUENCE   392 AA;   39888 MW; 82BBABDD99D6F567 CRC64;

Query Match   100.0%; Score 1969; DB 16; Length 392;
Best Local Similarity   100.0%; Pred. No. 7.9e-119;
Matches 392; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MSRAFIIDPTISAIDGLYDLLGTGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA   60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MSRAFIIDPTISAIDGLYDLLGTGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA   60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120

Qy   121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180

Qy   181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGVPGLT 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGVPGLT 240

Qy   241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300

Qy   301 STRQALRPRADGPVGAAAEQVGGGSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 STRQALRPRADGPVGAAAEQVGGGSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKY 360

Qy   361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
           |||||||||||||||||||||||||||||||
Db   361 SEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392

RESULT 2
Q49722
ID   Q49722   PRELIMINARY;   PRT;   394 AA.
AC   Q49722;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein ML0405.
GN   ML0405 OR B1620_C2_213 OR MLCL383.01.
OS   Mycobacterium leprae.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1769;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=TN;
RX   MEDLINE=21128732; PubMed=11234002;
RA   Cole S.T., Eiglmeier K., Parkhill J., James K.D., Thomson N.R.,
RA   Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA   Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R.M., Devlin K., Duthoy S., Feltwell T., Fraser A., Hamlin N.,
RA   Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA   Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA   Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA   Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA   Barrell B.G.;
RT   "Massive gene decay in the leprosy bacillus.";
RL   Nature 409:1007-1011(2001).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Smith D.R., Robison K.;
RL   Submitted (MAR-1994) to the EMBL/GenBank/DDBJ databases.

RN   [3]
RP   SEQUENCE OF 231-394 FROM N.A.
RA   Oliver K., Harris D., Parkhill J., Barrell B.G., Rajandream M.A.;
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 231-394 FROM N.A.
RX   MEDLINE=93188700; PubMed=8446027;
RA   Eiglmeier K., Honore N., Woods S.A., Caudron B., Cole S.T.;
RT   "Use of an ordered cosmid library to deduce the genomic organization
RT   of Mycobacterium leprae.";
RL   Mol. Microbiol. 7:197-206(1993).
DR   EMBL; AL583918; CAC29913.1; -.
DR   EMBL; U00015; AAC43223.1; -.
DR   EMBL; Z97179; CAB09940.1; -.
KW   Leproma; ML0405; -.
KW   Hypothetical protein; Transmembrane; Complete proteome.
FT   TRANSMEM   279   299   POTENTIAL.
SQ   SEQUENCE   394 AA;   40754 MW; D0E455C74ED7A10C CRC64;

Query Match   63.9%; Score 1258; DB 16; Length 394;
Best Local Similarity 62.7%; Pred. No. 4.4e-73;
Matches 247; Conservative   55; Mismatches   90; Indels   2; Gaps   1;

Qy   1 MSRAFIIDPTISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALEELAAAFPGDGWLGSAA   60
           ||||||||||||||||| |||:||||||||||||||||||||||:||||||||||||||||
Db   1 MSGAFIIDPTLKAIEAWHALLGIGVPNDGGVLYSSLSFFEKALEHLAAAFPGDGWLGSAA   60

Qy   61 DKYAGKNRNHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLT 120
           |||||:||||:||||||||:||||||:::||::|||:||||:|||||:|||||||||:|||
Db   61 DKYAGQNRKRVDIFQELAELDKELIELIHNQANSVQTTRGLLDGAKKALLFVRPVAIDLN 120

Qy   121 YIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAIAD 180
           ||| |:|::|::|::||:|::|::|::|||||:|||||:|:|:::||::||||||:::||||
Db   121 YIPLVGSVMSASIQACACAAAMAAVSGGLAYLLVQTAIHTAKFVALLARLLAHLLASAVAD 180

Qy   181 IISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWTGLFSRGWSNLESFFAGVPGLT 240
           :::|||:::|||:||||||||||||:|:|||:::|:|::|:|||:|:|::||||:||||:
Db   181 VVSDGVAIIKGIVDHLWHFIAGALITGLKDIVEKIIHWFFGLFSHWSRLHSFFGGIPGLS 240

Qy   241 GATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAA 300
           |||||||||||||||||:|||:|||||||:||::|||||||||||||||||||::|||||
Db   241 GATSGLSQVTGLFGVPGLAGSSGLLSGESLLSTENLPSLAGVGAGLGLGSSIPQLAQLHAA 300

Qy   301 STRQALRPRADGPVGAAAEQVGGGSQLVSAQGSQGMGGPVGMGGMHPSSGASK--GTTTK 358
           ||||||:||:||||::||||:||||||:|||||||||||||||||||||||||| |||:
Db   301 STRQGTRSQAGVSAELSTEQFGGQQEPVSAQGSQGMGGNTPASTKSKKDERKKK 360

Qy   359 KYSEGAAAGTEDAERAPVEADAGGGQKVLVRNVV 392
           |||||||||||||||||||||::|:|:|||
Db   361 KYSEGAAAGTDDAERAPIEVGSGGGKRALAQHVV 394

RESULT 3
P96213
ID   P96213   PRELIMINARY;   PRT;   402 AA.
AC   P96213;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT   01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE   Hypothetical protein Rv3864.
GN   RV3864 OR MTCY01A6.04C OR MT3978.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,

RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z83864; CAB06237.1; -.
DR   EMBL; AE007189; AAK48347.1; -.
DR   TIGR; MT3978; -.
DR   TubercuList; Rv3864; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE   402 AA;   42068 MW;   B40E0E22D482765 CRC64;

Query Match   19.3%;   Score 380.5;   DB 16;   Length 402;
Best Local Similarity   30.1%;   Pred. No. 1.3e-16;
Matches   129;   Conservative   47;   Mismatches 157;   Indels   95;   Gaps   13;

Qy   10 TISAIDGLYDLLGIGIPNQGGILYSSLEYFEKALELAAAFPGDGWLGSAADKYAGKNRN 69
              ::|::|||||||:::|:::::::::|||::::::::||::|::|:::::||
Db   11 TSNFIWGQLLLGEGIPDPGCIFNTGSSLFKQISDKMGLAIPGTNWIGQAAEAYLNQNIA 70

Qy   70 HVNFPFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLTYIPVVGHAL 129
             :|:|:::::|:|:|:|:::|::|::||:|::::::::|||:||
Db   71 QQLRAQVMGDLDKLTGNMISNQAKYVSDTRDVLRAMKKMIDGVYKVCKGLEKIPLLGHLW 130

Qy   130 SAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLLAELVAAAIADIISDVADII 189
             :::|:::::::|::::||:|::::||::|:|:|::|||||::|||||:::|::
Db   131 SWELAIPMSGIAMAVVGGALLYLITMTLMNATNL-------------------------- 164

Qy   190 KGTLGEVWEFITN-----ALNGLKELWDKLTG-W----------------VTGLFSRGW-- 226
              ||:||||:|| |::::|::||: :   :|:||||
Db   165 RGILGRLIEMLTTLPKFPGLPGLFSLPDIIDGLWPPKLPDIPIPGLPDPFKWPPT 224

Qy   227 ------SNLESF------------FAGVPGLTGATSGLSQVTGLF-GAAGL------------ 258
                   :||||: |||||:||::||:||||:
Db   225 PGSPLFPDLPSFPGPGFPGFPEPAIPGFP-ALPGLSPI1PNLFPGLPGLGDLLPGVVGDLGKL 283

Qy   259 -SASSGLAHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAA 317
              :|||:|||||:|||:::|||||:||||||||:|||::::::|::::||:||:|:|::|||
Db   284 PTWTELAALPDFLGGFAGLPSL-GFGNLLSFASLPTVGQVTATMGQLQQLVAAGGGPSQL 342

Qy   318 AEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKYSEGAAAGTEDAERAPVE 377
              :||:::::|::|::::::|:::::::::::::::::|:::::::::||:|:|:::::|||||
Db   343 ASMGSQQAQLISSQAQQG-----------GQQHATLVSDK----KEDEEGVA-----EAERAPID 387

Qy   378 ADAGGGQK 385
              ::
Db   388 AGTAASQR 395

RESULT 4
ID   O53631
ID   O53631;   PRELIMINARY;   PRT;   496 AA.
AC   O53631;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   PGRS-family protein.
GN   RV0109 OR MTV031.03C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; AL021926; CAA17303.1; -.
DR   TubercuList; Rv0109; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Complete proteome.
SQ   SEQUENCE   496 AA;   42289 MW;   A3F1AD215EC11786 CRC64;

Query Match   8.5%;   Score 168;   DB 16;   Length 496;
Best Local Similarity   25.5%;   Pred. No. 0.0082;
Matches   82;   Conservative   25;   Mismatches 113;   Indels   102;   Gaps   13;

Qy   115 VAVDLTYIPVVGHALS---------AAFQAPFCAGAMAVVG------------------ 146
              ||||||:||||||:| ::||:||::::::|
Db   11 VAAAATHLAGIGALSTANAAAAAP--TTALSVAGADEVSVLIAALFEAYAQEYQALSAQ 68

Qy   147 ------------GALAYLVVKTLINATQLLKLLAKLIAELVAAAIADIISDVADIIKGT 192
                          :::::::::::::|:::::::::::
Db   69 ALAFHDQFVQALMNGAVCYAAAET-ANATPLQAL-----QTVQQNVLTVVNAPTQALLGR 122

Qy   193 LGEVWEFITNALNGLKELWDKLTGWVTG----LFSRGWSNLESFFAGVPGLTGATSGLSQ 248
              ||:|::::::||:|::::|||||:   ::::::::::|::||||||:||||:|
Db   123 ----PIIGNGANGLPN----------TGQDGGPGGLLFGNGGN------GGSGGVDQ 159

Qy   249 VTGLFGAAGL--SASSGLAHADSLASSASLPA----LAGIGGGSGFGGLPSLAQVHAAS 301
              ||:||:||:   :|||:||||:|||||:|| |:|||||||::|:|:|:::::
Db   160 AGGNGGAAGLIGNGGSGGVGGPGIAGSAGGAGGLLFGNGGPGGAGGIGTTGD------ 214

Qy   302 TRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSSGASKGTTTKKYS 361
              |:::::::::::::::::::::::|:|:::::::::::::::::::|:::::::|:::::
Db   215 ------GGPGAGGNAIG----LFGSGGTGGMGGVGGVGGMGGVGNGGVGNGGTAGLFG 261

Qy   362 EGAAAGTEDAERAPVEADAGGG 383
              :|::::::::|::|:|:|:
Db   262 HGGAGGAGGIGSADGGLGGGGG 283

RESULT 5
ID   Q8VKR5
ID   Q8VKR5;   PRELIMINARY;   PRT;   533 AA.
AC   Q8VKR5;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   PE_PGRS family protein.
GN   MT0118.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,

```
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE006923; AAK44341.1; -.
DR   TIGR; MT0118; -.
DR   InterPro; IPR002952; Eggshell.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   PRINTS; PR01228; EGGSHELL.
DR   ProDom; PD001223; PE_region; 1.
SQ   SEQUENCE  533 AA;  46785 MW;  56F0115F01F4D4E5 CRC64;

Query Match             8.5%;  Score 168;  DB 16;  Length 533;
Best Local Similarity  26.2%;  Pred. No. 0.009;
Matches   88;  Conservative   24;  Mismatches 140;  Indels   84;  Gaps   14;

Qy     82 RQLISLIHDQANAVQTTRDILEGAKKGLEFVR-PVAVDLTYIPVVG----------------- 126
          ||:|:|                         :                :  ::
Db     35 RSLMSLLITSPATVAAAATHLAGIGSALSTANAAAAPTTALSVAGADEVSVLIAALFEA     94

Qy    127 -----HALSA--AFQQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAAAI 178
               |||:|   ::  |||  |  |::     |::||::||:||||:|||||:|||
Db     95 YAQEYQALSAQALAFHDQF--VQALNMGAVCYAAAET-ANATPLQAL-----QTVVQQNV 145

Qy    179 ADIISDVADIIKGTLGEVWEFITNALNGLKELWDKLTGWVTG----LFSRGWSNLESFFA 234
          |||:||||:||||:||||| || |:| |||||:||||:||      :|||||||||:
Db    146 LTVVNAPTQALLGR-----PIIGNGANGLPN-----TGQDGGPGGLLFGNGGN------- 188

Qy    235 GVPGLTGATSGLSQVTGLFGAAGL--SASSGLAHADSLASSASLPA-----LAGIGGGSG 287
          ||||||||||||||||||||||||  | |||||| || |:|| |     |||||||
Db    189 -----GSGGGVDQAGGNGGAAGLIGNGGSGVGGPGIAGSAGGAGGAGGLLFGNGGPGG 242

Qy    288 FGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGGSQLVSAQGSQGMGGPVGMGGMHP 347
          ||| ||     : ||  |     |  ::|    ||| ||:|||||| ||:| |
Db    243 AGGIGTTGD----------GGPGGAGGNAIG----LFGSGGTGMGGVGGMGGVGN 284

Qy    348 SSGASKGTTTKKYSEGAAAGTEDAERAPVEADAGGG 383
          |||||||||||||||||||||| |  |||  | |||
Db    285 GGNAGNGGTAGLFGHGGAGGAGGIGSADGGLGGGGG 320

RESULT 6
Q8FND6
ID   Q8FND6            PRELIMINARY;     PRT;   536 AA.
AC   Q8FND6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein.
GN   CE2208.
OS   Corynebacterium efficiens.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Corynebacteriaceae; Corynebacterium.
OX   NCBI_TaxID=152794;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YS-314 / AJ 12310 / DSM 44549 / JCM 11189;
RA   Kawarabayasi Y., Yamazaki J., Hino Y., Kikuchi H., Nakamura Y.,
RA   Ikeo K., Suzuki M., Mashima J., Itoh T., Yamagishi A., Nishio Y.,
RA   Usuda Y., Sugimoto S.;
RT   "The entire genomic sequence of Corynebacterium efficiens YS-314.";
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP005221; BAC19018.1; -.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE  536 AA;  52889 MW;  A6291609CCCB2114 CRC64;

Query Match             7.9%;  Score 156.5;  DB 16;  Length 536;
Best Local Similarity  23.3%;  Pred. No. 0.05;
Matches   99;  Conservative   48;  Mismatches 163;  Indels 115;  Gaps   20;

Qy     42 ALELAAAFPG--DGWLGSAADKYAGKNRNHVNFFQEL-ADLDRQLISLIHDQANAV-Q  96
```

```
Db    126 SLDQLCSDFSGINSGAVSAAQSWT----TMASTISEVSASLDRVAGELLASNAGEVFEQ 181
          :|::|:|                 :               ||| |         ||||

Qy     97 TTRDILEGAKKG------LEFVRPVAVDLTYIPVVGHALSAAFQAPFCAGAMAVVGGAL 149
                         :  ::|     |:||:||:|  |    |::
Db    182 ASARIAEVAEAGSVFSQNAREMSRSVGT-LNQI-YMGHQMS-VFMA---AASIAMKEPA 235

Qy    150 AYLVVKTLINATQLLKLLAKLAELVAAAIADIISDVADI-----IKGTLGEVWEFITNAL 204
          |  |         |:||    |    |    |::     ||:|  |||:|||:||:|||
Db    236 ERAAAES--------AYLASFQSAFQGDVRAGVPGIDNLMRVKGADSGSGGL---AL 281

Qy    205 NGLKELWDKLTGWVT-GLFSRGWS----------------NLESFFAG------ 235
          ||||:||||||||| ||||||||                |||||||
Db    282 -GMSDIAGSGAGFTTHGLTPQGFSAAGGGGTAAAHSVGAGDFGAVADNLGGLDVGDLQTS 340

Qy    236 VPGLTGATSGLSQVTGLFGAAGLSASSG----------------------------- 263
          ||||||||||||||||||||||| :||
Db    341 TASVAGAGSTILSNASGLSGVGGMNASAGGGAMAAAPFMGMNGRHGPTAGQRFSNAGLNPM 400

Qy    264 ---LAHADSLASSASLPALAGIGGGSGFGGLPSL-AQVHAASTRQALRPRADGPVGAAAE 319
             |:|||  |  |::|||||||||| |||| |  |:|  ||  |:|  ||  ||:  :
Db    401 NARTQSASAMGPMGSMLGSAGAGGGAGVGGVGGAAGGNTHRGGMRQALAPQTGNPASAAGA 460

Qy    320 QV-GGGSQLVSAQGSQGMGGPV------GMGGMHPSSGASKGTTTKKYSEGAAAGTEDA 371
          |: |||||:|:|  :| |||        ||||| |||||||||||||||||||||| |
Db    461 QTPSGMMTGGVGKASQSTARPMMPMPMGMPGGAGGAQPKNTGKVKTVTSAVEQDANLAALLG 520

Qy    372 ERAPV 376
          |||:|
Db    521 DRGPV 525

RESULT 7
O06810
ID   O06810            PRELIMINARY;     PRT;  1329 AA.
AC   O06810;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein Rv1450c.
GN   Rv1450c OR MTCY493.04.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton J., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393;537-544(1998).
DR   EMBL; Z95844; CAB09271.1; -.
DR   TubercuList; Rv1450c; -.
DR   InterPro; IPR000084; PE_region.
DR   InterPro; IPR002173; PfkB.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; 2.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE  1329 AA;  107410 MW;  777A125F6DBAB234 CRC64;

Query Match             7.9%;  Score 155;  DB 16;  Length 1329;
Best Local Similarity  26.3%;  Pred. No. 0.22;
Matches   79;  Conservative   27;  Mismatches 132;  Indels   62;  Gaps   13;
```

Db   181 AXGGAGGSGGW--LLGNGVGGAGG-------------------QSL------LGGATGGAGGNA 218

Qy   326 QLVSAQGSQGMGGPVGMGGMHPSSGAS--KGTTTKKYSEGAAAGTEDAERAPVEADAGGG 383

Db   219 GLFGVGGTXGPGPGGPGGPGGVGGTGEGGGLGGTLYGAGGHGGAGGP-----GPIGGVGGHG 273

RESULT 9
Q8VK15
ID   Q8VK15   PRELIMINARY;   PRT;   738 AA.
AC   Q8VK15;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   PE_PGRS family protein.
GN   MT1499.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE007019; AAK45763.1; -.
DR   TIGR; MT1499; -.
DR   InterPro; IPR002952; Eggshell.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   PRINTS; PR01228; EGGSHELL.
DR   ProDom; PD001223; PE_region; 1.
SQ   SEQUENCE   738 AA; 60853 MW; A6F4645310B9D0D0 CRC64;

Query Match   7.7%; Score 151; DB 16; Length 738;
Best Local Similarity 26.6%; Pred. No. 0.17;
Matches   80; Conservative   25; Mismatches 132; Indels   64; Gaps   13;

Qy   110 EFVRPVAVDLTYI-PVVGHALSAAFQAPFCAGAMAVV--GGALAYLVVKTLIN------- 159
              :||:|: |::|:| | :||||
Db   10 ETVAAAASDVARIGSSIGVANSAA-----------AGSTTSVLAAGADEVSAAIATLFGSHAREY 63

Qy   160 --ATQLLKLLAKLAELVAAAIADIISDVADIIKGTLGEVWEFITNALN------GLKE 209
               :|:|| ||| |:| || | ||||
Db   64 QAISTQVAAFHDRFAQTLSAAVGSYVSAEA-TNAAPLATLEHNVLNALNAPTQALLGRPL 122

Qy   210 LWDKL----TGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGL 264
            || ||||||| || || ||||||| ||
Db   123 IGDGAAGAPGTGQAGGAGGILWGNGGAGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQ 179

Qy   265 AHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQ 324

Db   180 GAAGGAGGSGGW--LLGNGVGGAGG----------------QSL------LGGATGGAGGN 217

Qy   325 SQLVSAQGSQGMGGPVGMGGMHPSSGAS--KGTTTKKYSEGAAAGTEDAERAPVEADAGG 382

Db   218 AGLFGVGGTGGPGGPGGPGGVGGTGGAGGLGGTLYGAGGHGGAGGP-----GPIGGVGGH 272

Qy   383 G 383

Db   273 G 273

RESULT 10
O06808
ID   O06808   PRELIMINARY;   PRT;   741 AA.
AC   O06808;

Qy   110 EFVRPVAVDLTYI-PVVGHALSAAFQAPFCAGAMAVV--GGALAYLVVKTLIN------- 159
              :||:|: |::|:| | :||||
Db   9 ETVAAAALDVARI---GSSIGAANA--AAGSTTSVLAAGADEVSAAIATLFGSHAREYQ 63

Qy   160 --ATQLLKLLAKLAELVAAAIADIISDVADIIKGTLGEVWEFITNALN------GLKEL 210
               :|:|| ||| |:| || | ||||
Db   64 AISTQVAAFHDRFAQTLSAAVGSYVSAEA-TNAAPLATLEHNVLNALNAPTQALLGRPLI 122

Qy   211 WDKL----TGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLA 265
          || ||||||| || || ||||||| ||
Db   123 GDGAAGAPGTGQAGGAGGILWGNGGAGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQS 179

Qy   266 HADSLASSASLPALAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQS 325

Db   180 AAGGAGGSGGW--LLGNGVGGAGG-------------------QSL------LGGATGGAGGN 217

Qy   326 QLVSAQGSQGMGGPVGMGGMHPSSGAS--KGTTTKKYSEGAAAGTEDAERAPVEADAGGG 383

Db   218 GLFGVGGTGGPGGPGGPGGVGGTGEGGGLGGTLYGAGGHGGAGGP-----GPIGGVGGHG 272

RESULT 8
Q8VK17
ID   Q8VK17   PRELIMINARY;   PRT;   1408 AA.
AC   Q8VK17;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   PE_PGRS family protein.
GN   MT1497.1.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE007019; AAK45760.1; -.
DR   TIGR; MT1497; -.
DR   InterPro; IPR000084; PE_region.
DR   InterPro; IPR002173; PFKB.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; 2.
SQ   SEQUENCE   1408 AA; 113502 MW; EF472A34966EA6DC CRC64;

Query Match   7.7%; Score 152; DB 16; Length 1408;
Best Local Similarity 26.0%; Pred. No. 0.37;
Matches   78; Conservative   27; Mismatches 133; Indels   62; Gaps   13;

Qy   110 EFVRPVAVDLTYI-PVVGHALSAAFQAPFCAGAMAVV--GGALAYLVVKTLIN------- 159
              :||:|: |::|:| | :||||
Db   9 ETVAAAALDVARI---GSSIGAANA--AAGSTTSVLAAGADEVSAAIATLFGSHAREYQ 63

Qy   160 --ATQLLKLLAKLAELVAAAIADIISDVADIIKGTLGEVWEFITNALN------GLKEL 210
               :|:|| ||| |:| || | ||||
Db   65 AISTQVAAFHDRFAQTLSAAVGSYVSAEA-TNAAPLATLEHNVLNALNAPTQALLGRPLI 123

Qy   211 WDKL----TGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLA 265
          || ||||||| || || ||||||| ||
Db   124 GDGAAGAPGTGQAGGAGGILWGNGGAGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQS 180

Qy   266 HADSLASSASLPALAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQS 325

```
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   Hypothetical protein Rv1452c.
GN   RV1452C OR MTCY493.02.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
DR   EMBL; Z95844; CAB09269.1; -.
DR   TuberculIst; Rv1452c; -.
DR   InterPro; IPR000084; PE_region.
DR   Pfam; PF00934; PE; 1.
DR   ProDom; PD001223; PE_region; 1.
KW   Hypothetical protein; Complete proteome.
SQ   SEQUENCE 741 AA; 61087 MW; E9BD6795BD543C9B CRC64;

Query Match          7.7%;  Score 151; DB 16;  Length 741;
Best Local Similarity   26.6%;  Pred. No. 0.18;
Matches  80;  Conservative  25; Mismatches 132; Indels  64;  Gaps  13;

Qy      110 EFVRPVAVDLTYI-PVVGHALSAAFQAPFCAGAMAVV--GGALAYLVVKTLIN------- 159
                ||||   |        ||||                    |||
Db        9 ETVAAAASDVARIGSSIGVANSAA------AGSTTSVLAAGADEVSAAIATLFGSHAREY 62

Qy      160 ---ATQLLKILAKLAELVAAAIADIISDVADIIKGTLGEVWEPITNALN------GLKE 209
                 ||||   |  |  ||:||:       ||||
Db       63 QAISTQVAAFHDRPAQTLSAAVGSYVSAEA-TNAAPLATLEHRVLNALNAPTQALLGRPL 121

Qy      210 LWDKL-----TGWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGL 264
                            |||||||||||||||  |||||||
Db      122 IGDGAAGAPGTGQAGGAGGILWGNGGAGGSGAPGQVGGAGG--AAGLFGTGGAGGAGGA 178

Qy      265 AHADSLASSASLPALAGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQ 324
                                   |||    ||
Db      179 GAAGGAGGSGGW--LLGNGGVGGAGG---------QSL------LGGATGGGAGGN 216

Qy      325 SQLVSAQGSQGMGGPVGMGGMHPSSGAS--KGTTTKKYSEGAAAGTEDAERAPVEADAGG 382
                |||                              ||:
Db      217 AGLFGVGGTGGPGPGGPGGPGVGGTGGAGGLGGTLYGAGHGGAGGP----GPIGGVGGH 271

Qy      383 G 383
              |
Db      272 G 272
```

```
RESULT 11
O53775
ID   O53775;
AC   O53775;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   PGRS-family protein (PE PGRS family protein).
GN   RV0578C OR MT0607 OR MTV039.16C.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
```

```
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL021942; CAA17449.1; -.
DR   EMBL; AE006958; AAK44830.1; -.
DR   TIGR; MT0607; -.
DR   TuberculIst; Rv0578c; -.
DR   InterPro; IPR002952; Eggshell.
DR   InterPro; IPR000209; Peptidase_S8.
DR   InterPro; IPR000084; PE_region.
DR   InterPro; IPR002173; PfkB.
DR   Pfam; PF00934; PE; 1.
DR   PRINTS; PR01228; EGGSHELL.
DR   ProDom; PD001223; PE_region; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; 2.
DR   PROSITE; PS00138; SUBTILASE_SER; 1.
KW   Complete proteome.
FT   CONFLICT    363    363       G -> R (IN REF. 2).
SQ   SEQUENCE 1306 AA; 105964 MW; 843A30955FFA56B6 CRC64;

Query Match          7.6%;  Score 150; DB 16;  Length 1306;
Best Local Similarity   28.3%;  Pred. No. 0.44;
Matches  60;  Conservative  10; Mismatches  82;  Indels  60;  Gaps   8;

Qy      216 GWVTGLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSAS 275
                ||||||||||||||||:||        |||::||:|||||||    :||
Db      185 GWLYGNGGAGG----FGGA-GAVGGNGGAGGTAGLFGVGGAGGAGGNGIAGVTGTSAS 237

Qy      276 LP------ALAGIGGGSGFGGL-----------PSLAQVHAASTRQALRPRA 310
                            ||  | |||
Db      238 TPGGSGTGAGGIGGNGGAGGAGGVLMGNGGNGGAGGEGGPGGAGGASGAHATNLGA 297

Qy      311 DGPVGAAAEQVGGQSGQLVSAQGSQGWGGP-----VGMGGWH--------PSSG 350
                 ||                 ||:||:||         :|::|            
Db      298 DG------QAGGNGGNGGAGGTGVGGPGGGHGLLGLGGSHGAGGAGGSGGDGGAPGDG 350

Qy      351 ASKGTTTKKYSEGAAAGTEDAERAPVEADAGG 382
                   | :||                ||||
Db      351 GNGATGTWGHNLG-AGGTGGNGGNPGAGGAGG 381
```

```
RESULT 12
O80116
ID   O80116;
AC   O80116;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   L16.
```

PRELIMINARY;   PRT;   1306 AA.

PRELIMINARY;   PRT;   916 AA.

GN   L16.
OS   Bacteriophage bIL170.
OC   Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae.
OX   NCBI_TaxID=63118;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Crutz-Le Coq A.M., Cesselin B., Commissaire J., Anba J.,
RA   Kyriakidis S., Chopin M.C.;
RT   "Sequence and organization of the lactococcal isometric bIL170 phage
RT   genome.";
RL   Submitted (JUN-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF009630; AAC27195.1; -.
SQ   SEQUENCE   916 AA;   96586 MW;   101058AE212E195F CRC64;

Query Match   7.6%;   Score 149.5; DB 9; Length 916;
Best Local Similarity   21.8%; Pred. No. 0.29;
Matches   89; Conservative   58; Mismatches 144; Indels 117; Gaps   18;

Qy   1 MSRAFIIDPTISA---IDGLYDLLGIGIPNQGGILYSSLEYFE---KALEE-------- 45
           |: ||||| :   : | :| : | :| : ||
Db   357 LSQALI--PLEALTPVINALIDKMDDWGKGAGKAIENVKYFQDLFKQLQQNGAITQFS 414

Qy   46 -----LAAAFPG-DGWLGSAADKYAG------KN---RNHVNFFQELADLRQLISLIH 89
               ||| | :| | : : : |:   : :| :| | ::||
Db   415 AIWDNLKSAFGSVIGIIGNLIKSFAGVDESTSKNSTSVENVANTISSLANKFADITKKIA 474

Qy   90 DQANAVQTTRDILEGAKKGLEFVRPVAVDL--------TYIPVV-GHALSAAFQAP 136
          | :||| :   : | ::   : : || : ||| :||
Db   475 DFIGKISESKSAMDAIKVALVALAGAFVAMKVINGIIKAYETYNKIVEAGTIIQGAFNAI 534

Qy   137 FCAGAMAVVGGALAYLVVKTLINATQL--------------------IA 179
          :|| | | :| : : : : : || |
Db   535 MAVNPFVLLGIAIAAVVAGLVYFFTQTETGKKAWASFVDFLKSAWDGIVSFFSGIGQWFA 594

Qy   180 DIISDVADIIKGT----------------LGEVWEFI---------TNALNGLKELWDKLTGW 217
          |: ||| ||| ||: ||: | : | : |
Db   595 DIWNGAVDGAKGIWQGLVDWFIGIVQGIQNIWNGITTEFTTLWTTVIGGIQSVWGGVTGF 654

Qy   218 VTGLF-------------------------SRGWSNLESFPAGVPGLTG---ATSGLSQVTGLFGA 255
          :||: | :| || | :||||| | :|||||||||
Db   655 FSGIFDAVSSVVSTVFSAIGGFASSAWNVLVSVWSAVAGFFGGIFNAVSGV--VSSVFSA 712

Qy   256 AGLSASSGLAHADSLASSAS------LPALAGIGGG--SGFGGLPSLA 295
          |||||||||||||| ||||| |||| |||| ||||| |||
Db   713 IGSFASSAWGVVQSIWSAVSGFFSGIFNSVSSVVSGVFSALGGFASNA 760

RESULT 13
Q8VJ19
ID   Q8VJ19;   PRELIMINARY;   PRT;   628 AA.
AC   Q8VJ19;
DT   01-MAR-2002 (TrEMBLrel. 20, Created)
DT   01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   PE PGRS family protein.
GN   MT3476.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE007154; AAK47814.1; -.
DR   TIGR; MT3476; -.

DR   InterPro; IPR002952; Eggshell.
DR   InterPro; IPR000084; PE-region.
DR   InterPro; IPR002173; PFKB.
DR   Pfam; PF00934; PE; 1.
DR   PRINTS; PR01228; EGGSHELL.
DR   ProDom; PD001223; PE_region; 1.
DR   PROSITE; PS00583; PFKB_KINASES_1; 1.
SQ   SEQUENCE   628 AA;   53828 MW;   9F62B1E7CC2C567E CRC64;

Query Match   7.6%;   Score 149; DB 16; Length 628;
Best Local Similarity   24.7%; Pred. No. 0.19;
Matches   79; Conservative   29; Mismatches 122; Indels   90; Gaps   13;

Qy   117 VDLTYIPVVGHALSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLAELVAA 176
           | ||| | | | : | | |||||| | : || | :||| : :||
Db   36 VVMSFVVAVPEALARA-----ASDVANIGSALSAANAAAAGTT--GLLAAGADEVSA 86

Qy   177 AIADIISDVA--------------------DIIKGTLG----------EVWEFITNALN----- 205
          | ||||| | ||||| | |: ||| ||
Db   87 ALASLSGHAVSYQQVAQATALHDQFVQALTGAGGSYALTEAANVQQNLLNAINAFTQA 146

Qy   206 --GLKELWDKLTGWTG--------------LFSRGWSNLESFPAGVPGLTGATSGLSQVTGLFGA 255
            | | : | | | ::| :| || | :|||| |||: :||||||
Db   147 LLGRPLIGGAVGTASSPDGQDGGLLFGNGGAGYNS--AATPGMAGGNGGNA--GLIGN 201

Qy   256 AGLSASSGLAHADSLASSASLPALAGIGGGSFGGLPSLAQVHAASTRQALPRPRADGPVG 315
          ||||||||| ||||||| ||| |||:|||| | | : :| : | | :| | |
Db   202 GGTGSGGAGAAGGAGGSGGW--LYGNGNGIGG------------NAIVAGGAGGNG 246

Qy   316 AAAEQVGGQSQLVSAQGSQGMGGPGVMGGMHPSGASKGTTTKKYSEGAAA------- 366
          | | | |:|:: | | | :| : | : : |:
Db   247 GA----GGAAGLWGSGSGSGGQGG---NGLTGNDGVNPAPVTNPALNGAAGDSNIEPQTS 298

Qy   367 ---GTEDAERAPVEADAGGG 383
                | || | | ||
Db   299 VLIGTQGGGDGTPGGAGVNGG 318

RESULT 14
O53844
ID   O53844;   PRELIMINARY;   PRT;   749 AA.
AC   O53844;
DT   01-JUN-1998 (TrEMBLrel. 06, Created)
DT   01-JUN-1998 (TrEMBLrel. 06, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   PGRS-family protein (PE_PGRS family protein).
GN   RV0833 OR MT0854.1 OR MTV043.25.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton J., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393.537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,

RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AL022004; CAA17639.1; -.
DR   EMBL; AE006974; AAK45096.1; -.
DR   TIGR; MT0854.1; -.
DR   Tuberculist; Rv0833; -.
KW   Complete proteome.
FT   CONFLICT   514   514   G -> A (IN REF. 2).
FT   CONFLICT   584   584   S -> G (IN REF. 2).
SQ   SEQUENCE   749 AA;   57448 MW;   C3C91307F9A8A2EC CRC64;

Query Match   7.5%;   Score 148.5;   DB 16;   Length 749;
Best Local Similarity   31.2%;   Pred. No. 0.26;
Matches 53;   Conservative 11;   Mismatches 83;   Indels 23;   Gaps   4;

Qy   220 GLFSRGWSNLESFFAGVPGLTGATSGLSQVTGLFGAAGLSASSGLAHADSLASSASLPAL 279
          |: :: ||| ||||| |||:   :: | |||||| || ||| || ||
Db   27   GVGGAGGDSAVAGVTGGAGGRAALLFGAGGAGGSGGSGAAGGAGGAGGAGGAGGAGGL   86

Qy   280 AGIGGGSGFGGLPSLAQVHAASTRQALRPRADGPVGA--AAEQVGGQSQLVSAQGSCQMG 337
          : || | : :: ||| | |||| || ||| :
Db   87   PASGSGGFGGFGASTGTGGAGGT-------GGAGGLFASGGVGGVGGTGGGAGSGGTGCVG 137

Qy   338 GPVGMGGMHPSSGA-----SKGTTTKKYSEGAAAGTEDAERAPVEADAGG 382
          ||| || || | | ||| :||||||| |||||
Db   138 GTGGAGGLFASGGAGGAGGSGGT-------GGAGGTGGAGGGLFGAGGAGG 180

RESULT 15
P95182
ID   P95182;   PRELIMINARY;   PRT;   409 AA.
AC   P95182;
DT   01-MAY-1997 (TrEMBLrel. 03, Created)
DT   01-AUG-1993 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   RV3144C (PPE family protein).
GN   RV3144C OR MT3231 OR MTCY03A2.14.
OS   Mycobacterium tuberculosis.
OC   Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC   Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX   NCBI_TaxID=1773;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=H37Rv;
RX   MEDLINE=98295987; PubMed=9634230;
RA   Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA   Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekaia F.,
RA   Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA   Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA   Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA   Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA   Rutter S., Seeger K., Skelton J., Squares S., Squares R.,
RA   Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT   "Deciphering the biology of Mycobacterium tuberculosis from the
RT   complete genome sequence.";
RL   Nature 393:537-544(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CDC 1551 / Oshkosh;
RA   Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA   Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
RA   Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
RA   Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA   Bishai W.;
RT   "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT   laboratory strains.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; Z83867; CAB06293.1; -.
DR   EMBL; AE007137; AAK47571.1; -.
DR   TIGR; MT3231; -.
DR   Tuberculist; Rv3144c; -.

DR   InterPro; IPR000030; Microbac_PPE.
DR   Pfam; PF00823; PPE; 1.
KW   Complete proteome.
FT   CONFLICT   226   226   S -> G (IN REF. 2).
SQ   SEQUENCE   409 AA;   39098 MW;   1D0C648337785A2 CRC64;

Query Match   7.5%;   Score 148;   DB 16;   Length 409;
Best Local Similarity   24.2%;   Pred. No. 0.12;
Matches 94;   Conservative 36;   Mismatches 137;   Indels 122;   Gaps 15;

Qy   12   SAIDGLYDLLGIGIPNQGGI---LYSSLEYFEKALEELAAAFPGDGWLGSAADKYAGKNR   68
          : |||| || ::| ||| | :|| ||| :||| |||:| ||
Db   29   AAWDGLAEELGTAAQSFASVTAGLAGQAWQGPAALAMAAAAAPYAGWLTAAAAQSAG---   85

Qy   69   NHVNFFQELADLDRQLISLIHDQANAVQTTRDILEGAKKGLEFVRPVAVDLTYIPVVGHA 128
          ::|: :   :| :: |:
Db   86   ----------------------AAGQARAVAS---IFEAAQAA---------TVLP---- 107

Qy   129 LSAAFQAPFCAGAMAVVGGALAYLVVKTLINATQLLKLLAKLLAELVAAAIADIISDVADI 188
            : : :   :| :||| : |
Db   108 ------------AAVAANRDAFVQLV-------------MTNLFGQNAPLIAAA------ 136

Qy   189 IKGTLGEVWEFITNALNGLKELWDKLTGWVTGLFSRGWSNLESFFAGV-PGLTGATSGLS 247
          :|: | : | :||| :|||
Db   137 -EGVYEEMWAADVAAMSGYYSGASAIAAQVV-----PWASLLQRFPGLGAGATGATGGES 190

Qy   248 QVTGL--------FGAAGLSAS-SGLAHADSLASSASL---PALAGIGGGSGFG 289
          |||:   : || : || :   :| :: |||| |||
Db   191 VGTGATGSGESVGTGGGESVGTGGATASGGGVGYVGSGVASAGLAAGDPAHGSVGGVQQNFGG 250

Qy   290 GLPSLAQVHAASTRQALRPRADGPVGAAAEQVGGQSQLVSAQGSQGMGGPVGMGGMHPSS 349
          ||| | | ::: :::::   :| | |||| | ||| | :
Db   251 GDVGAGDVVASSATSA-------HAGVVSPGFIGAPLALAALGQNARGG----------T 293

Qy   350 GASKGTTTKKYSEGAAAGTEDAERAPVEA 378
          :: |||: ||| |||| |
Db   294 NSAPGTATE----SARAPEPAASAPPEA 317

Search completed: November 5, 2003, 17:02:38
Job time : 59 secs